



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12N 15/12, C07K 14/72		A2	(11) International Publication Number: WO 00/31258
			(43) International Publication Date: 2 June 2000 (02.06.00)
(21) International Application Number:	PCT/US99/23687	09/416,760	12 October 1999 (12.10.99) US
(22) International Filing Date:	13 October 1999 (13.10.99)		
(30) Priority Data:			
60/109,213	20 November 1998 (20.11.98)	US	(71) Applicant (for all designated States except US): ARENA PHARMACEUTICALS, INC. [US/US]; 6166 Nancy Ridge Drive, San Diego, CA 92121 (US).
60/120,416	16 February 1999 (16.02.99)	US	(72) Inventors; and
60/121,852	26 February 1999 (26.02.99)	US	(75) Inventors/Applicants (for US only): CHEN, Ruoping [CN/US];
60/123,946	12 March 1999 (12.03.99)	US	5296 Timber Branch Way, San Diego, CA 92130 (US).
60/123,949	12 March 1999 (12.03.99)	US	DANG, Huong, T. [US/US]; 5352 Oak Park Drive, San
60/136,436	28 May 1999 (28.05.99)	US	Diego, CA 92105 (US). LIAW, Chen, W. [US/US]; 7668
60/136,437	28 May 1999 (28.05.99)	US	Salix Place, San Diego, CA 92129 (US). LIN, I-Lin [-/US];
60/136,439	28 May 1999 (28.05.99)	US	8291-7 Gold Coast Drive, San Diego, CA 92126 (US).
60/136,567	28 May 1999 (28.05.99)	US	
60/137,127	28 May 1999 (28.05.99)	US	(74) Agents: MILLER, Suzanne, E. et al.; Woodcock Washburn
60/137,131	28 May 1999 (28.05.99)	US	Kurtz Mackiewicz & Norris LLP, 46th floor, One Liberty
60/141,448	29 June 1999 (29.06.99)	US	Place, Philadelphia, PA 19103 (US).
60/156,653	29 September 1999 (29.09.99)	US	
60/156,633	29 September 1999 (29.09.99)	US	(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG,
60/156,555	29 September 1999 (29.09.99)	US	BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE,
60/156,634	29 September 1999 (29.09.99)	US	ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP,
60/157,280	1 October 1999 (01.10.99)	US	KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA,
60/157,294	1 October 1999 (01.10.99)	US	MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU,
60/157,281	1 October 1999 (01.10.99)	US	SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG,
60/157,293	1 October 1999 (01.10.99)	US	US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE,
60/157,282	1 October 1999 (01.10.99)	US	LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM,
09/417,044	12 October 1999 (12.10.99)	US	AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT,
			BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU,
			MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM,
			GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: HUMAN ORPHAN G PROTEIN-COUPLED RECEPTORS

(57) Abstract

The invention disclosed in this patent document relates to transmembrane receptors, more particularly to endogenous, human orphan G protein-coupled receptors.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		

HUMAN ORPHAN G PROTEIN-COUPLED RECEPTORS

This patent document claims priority benefit of each of the following applications, all filed with the United States Patent and Trademark Office via U.S. Express Mail on the indicated filing dates: U.S. Provisional Number 60/121,852, filed; February 26, 1999 claiming the benefit of U.S. Provisional Number 60/109,213, filed November 20, 1998; U.S. Provisional Number 60/120,416, filed February 16, 1999; U.S. Provisional Number 60/123,946, filed March 12, 1999; U.S. Provisional Number 60/123,949, filed March 12, 1999; U.S. Provisional Number 60/136,436, filed May 28, 1999; U.S. Provisional Number 60/136,439, filed May 28, 1999; U.S. Provisional Number 60/136,567, filed May 28, 1999; U.S. Provisional Number 60/137,127, filed May 28, 1999; U.S. Provisional Number 60/137,131, filed May 28, 1999; U.S. Provisional Number 141,448, filed June 29, 1999 claiming priority from U.S. Provisional Number 60/136,437, filed May 28, 1999; U.S. Provisional Number _____ (Arena Pharmaceuticals, Inc. docket number CHN10-1), filed September 29, 1999; U.S. Provisional Number 60/156,333, filed September 29, 1999; U.S. Provisional Number 60/156,555, filed September 29, 1999; U.S. Provisional Number 60/156,634, filed September 29, 1999; U.S. Provisional Number _____ (Arena Pharmaceuticals, Inc. docket number RUP6-1), filed October 1, 1999; U.S. Provisional Number _____ (Arena Pharmaceuticals, Inc. docket number RUP7-1), filed October 1, 1999; U.S. Provisional Number _____ (Arena Pharmaceuticals, Inc. docket number CHN6-1), filed October 1, 1999; U.S. Provisional

- 2 -

Number _____ (Arena Pharmaceuticals, Inc. docket number RUP5-1), filed October 1, 1999; U.S. Provisional Number _____ (Arena Pharmaceuticals, Inc. docket number CHN9-1), filed October 1, 1999. This patent document is related to U.S. Serial Number 09/170,496 filed October 13, 1998, and U.S. Serial Number unknown (Woodcock 5 Washburn Kurtz Mackiewicz & Norris, LLP docket number AREN-0054) filed on October 12, 1999 (via U.S. Express Mail) both being incorporated herein by reference. This patent document also is related to U.S. Serial No. 09/364,425; filed July 30, 1999, which is incorporated by reference in its entirety. This application also claims priority to U.S. Serial Number _____ (Woodcock, Washburn, Kurtz, Makiewicz & Norris, LLP 10 docket number AREN-0050), filed on October 12, 1999 (via U.S. Express Mail), incorporated by reference herein in its entirety. Each of the foregoing applications are incorporated herein by reference in their entirety.

FIELD OF THE INVENTION

The invention disclosed in this patent document relates to transmembrane receptors, 15 and more particularly to endogenous, orphan, human G protein-coupled receptors ("GPCRs").

BACKGROUND OF THE INVENTION

Although a number of receptor classes exist in humans, by far the most abundant and therapeutically relevant is represented by the G protein-coupled receptor (GPCR or GPCRs) 20 class. It is estimated that there are some 100,000 genes within the human genome, and of these, approximately 2% or 2,000 genes, are estimated to code for GPCRs. Receptors, including GPCRs, for which the endogenous ligand has been identified are referred to as "known" receptors, while receptors for which the endogenous ligand has not been identified

- 3 -

are referred to as "orphan" receptors. GPCRs represent an important area for the development of pharmaceutical products: from approximately 20 of the 100 known GPCRs, 60% of all prescription pharmaceuticals have been developed. This distinction is not merely semantic, particularly in the case of GPCRs. Thus, the orphan GPCRs are to the pharmaceutical industry what gold was to California in the late 19th century – an opportunity to drive growth, expansion, enhancement and development.

GPCRs share a common structural motif. All these receptors have seven sequences of between 22 to 24 hydrophobic amino acids that form seven alpha helices, each of which spans the membrane (each span is identified by number, *i.e.*, transmembrane-1 (TM-1), transmembrane-2 (TM-2), etc.). The transmembrane helices are joined by strands of amino acids between transmembrane-2 and transmembrane-3, transmembrane-4 and transmembrane-5, and transmembrane-6 and transmembrane-7 on the exterior, or "extracellular" side, of the cell membrane (these are referred to as "extracellular" regions 1, 2 and 3 (EC-1, EC-2 and EC-3), respectively). The transmembrane helices are also joined by strands of amino acids between transmembrane-1 and transmembrane-2, transmembrane-3 and transmembrane-4, and transmembrane-5 and transmembrane-6 on the interior, or "intracellular" side, of the cell membrane (these are referred to as "intracellular" regions 1, 2 and 3 (IC-1, IC-2 and IC-3), respectively). The "carboxy" ("C") terminus of the receptor lies in the intracellular space within the cell, and the "amino" ("N") terminus of the receptor lies in the extracellular space outside of the cell.

Generally, when an endogenous ligand binds with the receptor (often referred to as "activation" of the receptor), there is a change in the conformation of the intracellular region that allows for coupling between the intracellular region and an intracellular "G-protein." It

- 4 -

has been reported that GPCRs are "promiscuous" with respect to G proteins, *i.e.*, that a GPCR can interact with more than one G protein. *See*, Kenakin, T., 43 *Life Sciences* 1095 (1988). Although other G proteins exist, currently, Gq, Gs, Gi, and Go are G proteins that have been identified. Endogenous ligand-activated GPCR coupling with the G-protein 5 begins a signaling cascade process (referred to as "signal transduction"). Under normal conditions, signal transduction ultimately results in cellular activation or cellular inhibition. It is thought that the IC-3 loop as well as the carboxy terminus of the receptor interact with the G protein.

Under physiological conditions, GPCRs exist in the cell membrane in equilibrium 10 between two different conformations: an "inactive" state and an "active" state. A receptor in an inactive state is unable to link to the intracellular signaling transduction pathway to produce a biological response. Changing the receptor conformation to the active state allows linkage to the transduction pathway (via the G-protein) and produces a biological response. A receptor may be stabilized in an active state by an endogenous ligand or a compound such 15 as a drug.

SUMMARY OF THE INVENTION

Disclosed herein are human endogenous orphan G protein-coupled receptors.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B provide reference "grids" for certain dot-blots provided herein 20 (*see also*, Figure 2A and 2B, respectively).

Figures 2A and 2B provide reproductions of the results of certain dot-blot analyses resulting from hCHN3 and hCHN8, respectively (*see also*, Figures 1A and 1B, respectively).

Figure 3 provides a reproduction of the results of RT-PCR analysis of hRUP3.

- 5 -

Figure 4 provides a reproduction of the results of RT-PCR analysis of hRUP4.

Figure 5 provides a reproduction of the results of RT-PCR analysis of hRUP6.

DETAILED DESCRIPTION

The scientific literature that has evolved around receptors has adopted a number of 5 terms to refer to ligands having various effects on receptors. For clarity and consistency, the following definitions will be used throughout this patent document. To the extent that these definitions conflict with other definitions for these terms, the following definitions shall control:

AMINO ACID ABBREVIATIONS used herein are set out in Table 1:

10	TABLE 1	
	ALANINE	ALA
	ARGININE	ARG
	ASPARAGINE	ASN
	ASPARTIC ACID	ASP
15	CYSTEINE	CYS
	GLUTAMIC ACID	GLU
	GLUTAMINE	GLN
	GLYCINE	GLY
	HISTIDINE	HIS
20	ISOLEUCINE	ILE
	LEUCINE	LEU
	LYSINE	LYS
	METHIONINE	MET
	PHENYLALANINE	PHE
25	PROLINE	PRO
	SERINE	SER
	THREONINE	THR
	TRYPTOPHAN	TRP
	TYROSINE	TYR
30	VALINE	VAL

COMPOSITION means a material comprising at least one component.

ENDOGENOUS shall mean a material that a mammal naturally produces.

ENDOGENOUS in reference to, for example and not limitation, the term "receptor," shall mean that which is naturally produced by a mammal (for example, and not limitation, a

- 6 -

human) or a virus. By contrast, the term **NON-ENDOGENOUS** in this context shall mean that which is not naturally produced by a mammal (for example, and not limitation, a human) or a virus.

HOST CELL shall mean a cell capable of having a Plasmid and/or Vector 5 incorporated therein. In the case of a prokaryotic Host Cell, a Plasmid is typically replicated as a autonomous molecule as the Host Cell replicates (generally, the Plasmid is thereafter isolated for introduction into a eukaryotic Host Cell); in the case of a eukaryotic Host Cell, a Plasmid is integrated into the cellular DNA of the Host Cell such that when the eukaryotic Host Cell replicates, the Plasmid replicates. Preferably, for the purposes of the invention 10 disclosed herein, the Host Cell is eukaryotic, more preferably, mammalian, and most preferably selected from the group consisting of 293, 293T and COS-7 cells.

LIGAND shall mean an endogenous, naturally occurring molecule specific for an endogenous, naturally occurring receptor.

NON-ORPHAN RECEPTOR shall mean an endogenous naturally occurring 15 molecule specific for an endogenous naturally occurring ligand wherein the binding of a ligand to a receptor activates an intracellular signaling pathway.

ORPHAN RECEPTOR shall mean an endogenous receptor for which the endogenous ligand specific for that receptor has not been identified or is not known.

PLASMID shall mean the combination of a Vector and cDNA. Generally, a Plasmid 20 is introduced into a Host Cell for the purposes of replication and/or expression of the cDNA as a protein.

VECTOR in reference to cDNA shall mean a circular DNA capable of incorporating at least one cDNA and capable of incorporation into a Host Cell.

- 7 -

The order of the following sections is set forth for presentational efficiency and is not intended, nor should be construed, as a limitation on the disclosure or the claims to follow.

Identification of Human GPCRs

5 The efforts of the Human Genome project have led to the identification of a plethora of information regarding nucleic acid sequences located within the human genome; it has been the case in this endeavor that genetic sequence information has been made available without an understanding or recognition as to whether or not any particular genomic sequence does or may contain open-reading frame information that translate human proteins.

10 Several methods of identifying nucleic acid sequences within the human genome are within the purview of those having ordinary skill in the art. For example, and not limitation, a variety of GPCRs, disclosed herein, were discovered by reviewing the GenBank™ database, while other GPCRs were discovered by utilizing a nucleic acid sequence of a GPCR, previously sequenced, to conduct a BLAST™ search of the EST database. **Table A**, below, 15 lists the disclosed endogenous orphan GPCRs along with a GPCR's respective homologous GPCR:

TABLE A

	Disclosed	Accession	Open Reading	Per Cent	Reference To
	Human	Number	Frame	Homology	Homologous
20	Orphan	Identified	(Base Pairs)	To Designated	GPCR
				GPCR	(Accession No.)
	hARE-3	AL033379	1,260 bp	52.3% LPA-R	U92642
	hARE-4	AC006087	1,119 bp	36% P2Y5	AF000546

- 8 -

	hARE-5	AC006255	1,104 bp	32% <i>Oryzias</i>	D43633
	hGPR27	AA775870	1,128 bp	<i>latipes</i>	
	hARE-1	AI090920	999 bp	43%	D13626
5	hARE-2	AA359504	1,122 bp	KIAA0001	
	hPPR1	H67224	1,053 bp	53% GPR27	L31581
	hG2A	AA754702	1,113 bp	39% EBI1	L36148
	hRUP3	AL035423	1,005 bp	31% GPR4	2133653
				<i>Drosophila</i>	
	hRUP4	AI307658	1,296 bp	<i>melanogaster</i>	NP_004876
				32% pNPGPR	
				28% and 29 %	AAC41276
				<i>Zebrafish</i> Ya	and
				and Yb,	AAB94616
	hRUP5	AC005849	1,413 bp	respectively	
				25% DEZ	Q99788
10	hRUP6	AC005871	1,245 bp	23% FMLPR	P21462
	hRUP7	AC007922	1,173 bp	48% GPR66	NP_006047
	hCHN3	EST 36581	1,113 bp	43% H3R	AF140538
	hCHN4	AA804531	1,077 bp	53% GPR27	
	hCHN6	EST 2134670	1,503 bp	32% thrombin	4503637
15	hCHN8	EST 764455	1,029 bp	36% edg-1	NP_001391
				47%	D13626
	hCHN9	EST 1541536	1,077 bp	KIAA0001	
	hCHN10	EST 1365839	1,055 bp	41% LTB4R	NM_000752
				35% P2Y	NM_002563

Receptor homology is useful in terms of gaining an appreciation of a role of the disclosed receptors within the human body. Additionally, such homology can provide insight 20 as to possible endogenous ligand(s) that may be natural activators for the disclosed orphan GPCRs.

B. Receptor Screening

Techniques have become more readily available over the past few years for

- 9 -

endogenous-ligand identification (this, primarily, for the purpose of providing a means of conducting receptor-binding assays that require a receptor's endogenous ligand) because the traditional study of receptors has always proceeded from the a priori assumption (historically based) that the endogenous ligand must first be identified before discovery could proceed to 5 find antagonists and other molecules that could affect the receptor. Even in cases where an antagonist might have been known first, the search immediately extended to looking for the endogenous ligand. This mode of thinking has persisted in receptor research even after the discovery of constitutively activated receptors. What has not been heretofore recognized is that it is the active state of the receptor that is most useful for discovering agonists, partial 10 agonists, and inverse agonists of the receptor. For those diseases which result from an overly active receptor or an under-active receptor, what is desired in a therapeutic drug is a compound which acts to diminish the active state of a receptor or enhance the activity of the receptor, respectively, not necessarily a drug which is an antagonist to the endogenous ligand. This is because a compound that reduces or enhances the activity of the active receptor state 15 need not bind at the same site as the endogenous ligand. Thus, as taught by a method of this invention, any search for therapeutic compounds should start by screening compounds against the ligand-independent active state.

As is known in the art, GPCRs can be "active" in their endogenous state even without the binding of the receptor's endogenous ligand thereto. Such naturally-active receptors can 20 be screened for the direct identification (*i.e.*, without the need for the receptor's endogenous ligand) of, in particular, inverse agonists. Alternatively, the receptor can be "activated" via, *e.g.*, mutation of the receptor to establish a non-endogenous version of the receptor that is active in the absence of the receptor's endogenous ligand.

- 10 -

Screening candidate compounds against an endogenous or non-endogenous, constitutively activated version of the human orphan GPCRs disclosed herein can provide for the direct identification of candidate compounds which act at this cell surface receptor, without requiring use of the receptor's endogenous ligand. By determining areas within 5 the body where the endogenous version of human GPCRs disclosed herein is expressed and/or over-expressed, it is possible to determine related disease/disorder states which are associated with the expression and/or over-expression of the receptor; such an approach is disclosed in this patent document.

With respect to creation of a mutation that may evidence constitutive activation of 10 human orphan GPCRs disclosed herein is based upon the distance from the proline residue at which is presumed to be located within TM6 of the GPCR typically nears the TM6/IC3 interface (such proline residue appears to be quite conserved). By mutating the amino acid residue located 16 amino acid residues from this residue (presumably located in the IC3 region of the receptor) to, most preferably, a lysine residue, such activation may be obtained. 15 Other amino acid residues may be useful in the mutation at this position to achieve this objective.

C. Disease/Disorder Identification and/or Selection

Preferably, the DNA sequence of the human orphan GPCR can be used to make a probe for (a) dot-blot analysis against tissue-mRNA, and/or (b) RT-PCR identification of 20 the expression of the receptor in tissue samples. The presence of a receptor in a tissue source, or a diseased tissue, or the presence of the receptor at elevated concentrations in diseased tissue compared to a normal tissue, can be preferably utilized to identify a correlation with a treatment regimen, including but not limited to, a disease associated

- 11 -

with that disease. Receptors can equally well be localized to regions of organs by this technique. Based on the known functions of the specific tissues to which the receptor is localized, the putative functional role of the receptor can be deduced.

D. Screening of Candidate Compounds

5 1. Generic GPCR screening assay techniques

When a G protein receptor becomes constitutively active (i.e., active in the absence of endogenous ligand binding thereto), it binds to a G protein (e.g., Gq, Gs, Gi, Go) and stimulates the binding of GTP to the G protein. The G protein then acts as a GTPase and slowly hydrolyzes the GTP to GDP, whereby the receptor, under normal conditions, becomes 10 deactivated. However, constitutively activated receptors continue to exchange GDP to GTP.

A non-hydrolyzable analog of GTP, [³⁵S]GTP γ S, can be used to monitor enhanced binding to membranes which express constitutively activated receptors. It is reported that [³⁵S]GTP γ S can be used to monitor G protein coupling to membranes in the absence and presence of ligand. An example of this monitoring, among other examples well-known and 15 available to those in the art, was reported by Traynor and Nahorski in 1995. The preferred use of this assay system is for initial screening of candidate compounds because the system is generically applicable to all G protein-coupled receptors regardless of the particular G protein that interacts with the intracellular domain of the receptor.

2. Specific GPCR screening assay techniques

20 Once candidate compounds are identified using the "generic" G protein-coupled receptor assay (i.e., an assay to select compounds that are agonists, partial agonists, or inverse agonists), further screening to confirm that the compounds have interacted at the receptor site is preferred. For example, a compound identified by the "generic" assay may not bind to the

receptor, but may instead merely "uncouple" the G protein from the intracellular domain.

a. Gs and Gi.

Gs stimulates the enzyme adenylyl cyclase. Gi (and Go), on the other hand, inhibit this enzyme. Adenylyl cyclase catalyzes the conversion of ATP to cAMP; thus, 5 constitutively activated GPCRs that couple the Gs protein are associated with increased cellular levels of cAMP. On the other hand, constitutively activated GPCRs that couple the Gi (or Go) protein are associated with decreased cellular levels of cAMP. *See, generally,* "Indirect Mechanisms of Synaptic Transmission," Chpt. 8, From Neuron To Brain (3rd Ed.) Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992). Thus, assays that detect cAMP can 10 be utilized to determine if a candidate compound is, *e.g.*, an inverse agonist to the receptor (*i.e.*, such a compound would decrease the levels of cAMP). A variety of approaches known in the art for measuring cAMP can be utilized; a most preferred approach relies upon the use 15 of anti-cAMP antibodies in an ELISA-based format. Another type of assay that can be utilized is a whole cell second messenger reporter system assay. Promoters on genes drive expression of the proteins that a particular gene encodes. Cyclic AMP drives gene expression by promoting the binding of a cAMP-responsive DNA binding protein or transcription factor (CREB) which then binds to the promoter at specific sites called cAMP response elements and drives the expression of the gene. Reporter systems can be constructed which have a promoter containing multiple cAMP response elements before the reporter 20 gene, *e.g.*, β -galactosidase or luciferase. Thus, a constitutively activated Gs-linked receptor causes the accumulation of cAMP that then activates the gene and expression of the reporter protein. The reporter protein such as β -galactosidase or luciferase can then be detected using standard biochemical assays (Chen et al. 1995).

Go and Gq.

Gq and Go are associated with activation of the enzyme phospholipase C, which in turn hydrolyzes the phospholipid PIP₂, releasing two intracellular messengers: 5 diacycloglycerol (DAG) and inistol 1,4,5-triphosphate (IP₃). Increased accumulation of IP₃ is associated with activation of Gq- and Go-associated receptors. *See, generally, "Indirect Mechanisms of Synaptic Transmission," Chpt. 8, From Neuron To Brain (3rd Ed.) Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992).* Assays that detect IP₃ accumulation can be utilized to determine if a candidate compound is, *e.g.*, an inverse agonist to a Gq- or Go-associated receptor (*i.e.*, such a compound would decrease the levels of IP₃). Gq-associated receptors can also been examined using an AP1 reporter assay in that Gq-dependent phospholipase C causes activation of genes containing AP1 elements; thus, activated Gq-associated receptors will evidence an increase in the expression of such genes, whereby inverse agonists thereto will evidence a decrease in such expression, and agonists will 15 evidence an increase in such expression. Commercially available assays for such detection are available.

3. GPCR Fusion Protein

The use of an endogenous, constitutively activated orphan GPCR, or a non-endogenous, constitutively activated orphan GPCR, for screening of candidate compounds 20 for the direct identification of inverse agonists, agonists and partial agonists provides a unique challenge in that, by definition, the receptor is active even in the absence of an endogenous ligand bound thereto. Thus, it is often useful that an approach be utilized that can enhance the signal obtained by the activated receptor. A preferred approach is the use of a GPCR Fusion Protein.

- 14 -

Generally, once it is determined that a GPCR is or has been constitutively activated, using the assay techniques set forth above (as well as others), it is possible to determine the predominant G protein that couples with the endogenous GPCR. Coupling of the G protein to the GPCR provides a signaling pathway that can be assessed. Because it is most preferred that screening take place by use of a mammalian expression system, such a system will be expected to have endogenous G protein therein. Thus, by definition, in such a system, the constitutively activated orphan GPCR will continuously signal. In this regard, it is preferred that this signal be enhanced such that in the presence of, *e.g.*, an inverse agonist to the receptor, it is more likely that it will be able to more readily differentiate, particularly in the context of screening, between the receptor when it is contacted with the inverse agonist.

The GPCR Fusion Protein is intended to enhance the efficacy of G protein coupling with the GPCR. The GPCR Fusion Protein is preferred for screening with a non-endogenous, constitutively activated GPCR because such an approach increases the signal that is most preferably utilized in such screening techniques, although the GPCR Fusion Protein can also be (and preferably is) used with an endogenous, constitutively activated GPCR. This is important in facilitating a significant "signal to noise" ratio; such a significant ratio is import preferred for the screening of candidate compounds as disclosed herein.

The construction of a construct useful for expression of a GPCR Fusion Protein is within the purview of those having ordinary skill in the art. Commercially available expression vectors and systems offer a variety of approaches that can fit the particular needs of an investigator. The criteria of importance for such a GPCR Fusion Protein construct is that the GPCR sequence and the G protein sequence both be in-frame (preferably, the sequence for the GPCR is upstream of the G protein sequence) and that the "stop" codon of

- 15 -

the GPCR must be deleted or replaced such that upon expression of the GPCR, the G protein can also be expressed. The GPCR can be linked directly to the G protein, or there can be spacer residues between the two (preferably, no more than about 12, although this number can be readily ascertained by one of ordinary skill in the art). We have a preference (based upon convenience) of use of a spacer in that some restriction sites that are not used will, effectively, upon expression, become a spacer. Most preferably, the G protein that couples to the GPCR will have been identified prior to the creation of the GPCR Fusion Protein construct. Because there are only a few G proteins that have been identified, it is preferred that a construct comprising the sequence of the G protein (*i.e.*, a universal G protein construct) be available for insertion of an endogenous GPCR sequence therein; this provides for efficiency in the context of large-scale screening of a variety of different endogenous GPCRs having different sequences.

E. Other Utility

Although a preferred use of the human orphan GPCRs disclosed herein may be for the direct identification of candidate compounds as inverse agonists, agonists or partial agonists (preferably for use as pharmaceutical agents), these versions of human GPCRs can also be utilized in research settings. For example, *in vitro* and *in vivo* systems incorporating GPCRs can be utilized to further elucidate and understand the roles these receptors play in the human condition, both normal and diseased, as well as understanding the role of constitutive activation as it applies to understanding the signaling cascade. The value in human orphan GPCRs is that its utility as a research tool is enhanced in that by determining the location(s) of such receptors within the body, the GPCRs can be used to understand the role of these receptors in the human body before the endogenous ligand therefor is identified.

- 16 -

Other uses of the disclosed receptors will become apparent to those in the art based upon, *inter alia*, a review of this patent document.

EXAMPLES

The following examples are presented for purposes of elucidation, and not limitation, 5 of the present invention. While specific nucleic acid and amino acid sequences are disclosed herein, those of ordinary skill in the art are credited with the ability to make minor modifications to these sequences while achieving the same or substantially similar results reported below. Unless otherwise indicated below, all nucleic acid sequences for the disclosed endogenous orphan human GPCRs have been sequenced and verified. For 10 purposes of equivalent receptors, those of ordinary skill in the art will readily appreciate that conservative substitutions can be made to the disclosed sequences to obtain a functionally equivalent receptor.

Example 1 ENDOGENOUS HUMAN GPCRS

15 1. Identification of Human GPCRs

Several of the disclosed endogenous human GPCRs were identified based upon a review of the GenBank database information. While searching the database, the following cDNA clones were identified as evidenced below.

	Disclosed	Accession	Complete DNA	Open Reading	Nucleic Acid	Amino
20	Human	Number	Sequence	Frame	SEQ.ID.	Acid
	Orphan		(Base Pairs)	(Base Pairs)	NO.	SEQ.ID.
	GPCRs					NO.

- 17 -

hARE-3	AL033379	111,389 bp	1,260 bp	1	2
hARE-4	AC006087	226,925 bp	1,119 bp	3	4
hARE-5	AC006255	127,605 bp	1,104 bp	5	6
hRUP3	AL035423	140,094 bp	1,005 bp	7	8
5 hRUP5	AC005849	169,144 bp	1,413 bp	9	10
hRUP6	AC005871	218,807 bp	1,245 bp	11	12
hRUP7	AC007922	158,858 bp	1,173 bp	13	14

Other disclosed endogenous human GPCRs were identified by conducting a BLAST search of EST database (dbest) using the following EST clones as query sequences. The 10 following EST clones identified were then used as a probe to screen a human genomic library.

	Disclosed	Query	EST Clone/	Open	Nucleic Acid	Amino Acid
	Human	(Sequence)	Accession No.	Reading	SEQ.ID.NO.	SEQ.ID.NO.
	Orphan		Identified	Frame		
15	GPCRs			(Base Pairs)		
	hGPCR27	Mouse	AA775870	1,125 bp	15	16
	hARE-1	GPCR27 TDAG	1689643	999 bp	17	18
	hARE-2	GPCR27	AI090920 68530	1,122 bp	19	20
	hPPR1	Bovine	AA359504 238667	1,053 bp	21	22
20	hG2A	PPR1 Mouse	H67224 <i>See Example 2(a),</i> 1179426	1,113 bp	23	24
			<i>below</i>			

- 18 -

	hCHN3	N.A.	EST 36581	1,113 bp	25	26
	hCHN4	TDAG	(full length) 1184934	1,077 bp	27	28
	hCHN6	N.A.	AA804531 EST 2134670	1,503 bp	29	30
	hCHN8	KIAA0001	(full length) EST 764455	1,029 bp	31	32
5	hCHN 9	1365839	EST 1541536	1,077 bp	33	34
	hCHN10	Mouse EST	Human 1365839	1,005 bp	35	36
	hRUP4	1365839 N.A.	AI307658	1,296 bp	37	38

N.A. = "not applicable".

2. Full Length Cloning

10 a. hG2A (Seq. Id. Nos. 23 & 24)

Mouse EST clone 1179426 was used to obtain a human genomic clone containing all but three amino acid hG2A coding sequences. The 5'end of this coding sequence was obtained by using 5'RACE™, and the template for PCR was Clontech's Human Spleen Marathon-ready™ cDNA. The disclosed human G2A was amplified by PCR using the G2A 15 cDNA specific primers for the first and second round PCR as shown in SEQ.ID.NO.: 39 and SEQ.ID.NO.:40 as follows:

5'-CTGTGTACAGCAGTTCGCAGAGTG-3' (SEQ.ID.NO.: 39; 1st round PCR)

5'-GAGTGCAGGCAGAGCAGGTAGAC-3' (SEQ.ID.NO.: 40; second round PCR).

PCR was performed using Advantage™ GC Polymerase Kit (Clontech; manufacturing 20 instructions will be followed), at 94°C for 30 sec followed by 5 cycles of 94°C for 5 sec and 72°C for 4 min; and 30 cycles of 94° for 5 sec and 70° for 4 min. An approximate 1.3 Kb PCR fragment was purified from agarose gel, digested with Hind III and Xba I and cloned into the expression vector pRC/CMV2 (Invitrogen). The cloned-insert was sequenced using the T7 Sequenase™ kit (USB Amersham; manufacturer instructions will be followed) and

- 19 -

the sequence was compared with the presented sequence. Expression of the human G2A will be detected by probing an RNA dot blot (Clontech; manufacturer instructions will be followed) with the P³²-labeled fragment.

b. hCHN9 (Seq. Id. Nos. 33 & 34)

5 Sequencing of the EST clone 1541536 indicated that hCHN9 is a partial cDNA clone having only an initiation codon; *i.e.*, the termination codon was missing. When hCHN9 was used to "blast" against the data base (nr), the 3' sequence of hCHN9 was 100% homologous to the 5' untranslated region of the leukotriene B4 receptor cDNA, which contained a termination codon in the frame with hCHN9 coding sequence. To 10 determine whether the 5' untranslated region of LTB4R cDNA was the 3' sequence of hCHN9, PCR was performed using primers based upon the 5' sequence flanking the initiation codon found in hCHN9 and the 3' sequence around the termination codon found in the LTB4R 5' untranslated region. The 5' primer sequence utilized was as follows:

5'-CCCGAATTCCCTGCTTGCTCCCAGCTTGGCCC-3' (SEQ.ID.NO.: 41; sense) and

15 5'-TGTGGATCCTGCTGTCAAAGGTCCCATTCCGG-3' (SEQ.ID.NO.: 42; antisense).

PCR was performed using thymus cDNA as a template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 uM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 65°C for 1 min and 72 °C for 1 min and 10 sec. A 1.1kb fragment consistent with the predicted size was 20 obtained from PCR. This PCR fragment was subcloned into pCMV (*see* below) and sequenced (*see*, SEQ.ID.NO.: 33).

c. hRUP 4 (Seq. Id. Nos. 37 & 38)

The full length hRUP4 was cloned by RT-PCR with human brain cDNA (Clontech)

- 20 -

as templates:

5'-TCACAATGCTAGGTGTGGTC-3' (SEQ.ID.NO.: 43; sense) and
5'-TGCATAGACAATGGGATTACAG-3' (SEQ.ID.NO.: 44; antisense).

PCR was performed using TaqPlus™ Precision™ polymerase (Stratagene; manufacturing 5 instructions will be followed) by the following cycles: 94°C for 2 min; 94°C 30 sec; 55°C for 30 sec, 72°C for 45 sec, and 72°C for 10 min. Cycles 2 through 4 were repeated 30 times.

The PCR products were separated on a 1% agarose gel and a 500 bp PCR fragment was isolated and cloned into the pCRII-TOPO vector (Invitrogen) and sequenced using the 10 T7 DNA Sequenase™ kit (Amsham) and the SP6/T7 primers (Stratagene). Sequence analysis revealed that the PCR fragment was indeed an alternatively spliced form of AI307658 having a continuous open reading frame with similarity to other GPCRs. The completed sequence of this PCR fragment was as follows:

5'-TCACAATGCTAGGTGTGGCTGGCTGGCAGTCATCGTAGGATACCCATGTGGCAC
15 GTGCAACAACTTGAGATCAAATATGACTCCTATATGAAAAGGAACACATCTGCTGCTTAGAA
GAGTGGACCAGCCCTGTGCACCAGAAGATCTACACCACCTTCATCCTGTATCCTCTCC
TGCCTCTTATGGTGTGCTTATTCTGTACGTAAAATTGGTTATGAACTTGGATAAAAGAAAAGA
GTTGGGGATGGTTCACTGCTTCGAACATTACATGGAAAAGAAATGTCCAAAATAGCCAGGAAG
AAGAAACGAGCTGTCATTATGATGGTACAGTGGTGGCTCTTTGCTGTGCTGGCACCA
20 TTCCATGTTGCCATATGATGATTGAATACAGTAATTTGAAAAGGAATATGATGATGTCACA
ATCAAGATGATTTGCTATCGTGCAAATTATTGGATTTCCAACTCCATCTGTAATCCCATTG
TCTATGCA-3' (SEQ.ID.NO.: 45)

Based on the above sequence, two sense oligonucleotide primer sets:

5'-CTGCTTAGAAGAGTGGACCAG-3' (SEQ.ID.NO.: 46; oligo 1),

25 5'-CTGTGCACCAGAAGATCTACAC-3' (SEQ.ID.NO.: 47; oligo 2)

and two antisense oligonucleotide primer sets:

5'-CAAGGATGAAGGTGGTGTAGA-3' (SEQ.ID.NO.: 48; oligo 3)

5'-GTGTAGATCTTCTGGTGCACAGG-3' (SEQ.ID.NO.: 49; oligo 4)

were used for 3'- and 5'-race PCR with a human brain Marathon-Ready™ cDNA (Clontech,

- 21 -

Cat# 7400-1) as template, according to manufacture's instructions. DNA fragments generated by the RACE PCR were cloned into the pCRII-TOPO™ vector (Invitrogen) and sequenced using the SP6/T7 primers (Stratagene) and some internal primers. The 3' RACE product contained a poly(A) tail and a completed open reading frame ending at a TAA stop 5 codon. The 5' RACE product contained an incomplete 5' end; *i.e.*, the ATG initiation codon was not present.

Based on the new 5' sequence, oligo 3 and the following primer:

5'-GCAATGCAGGTCAAGTGAGC -3' (SEQ.ID.NO.: 50; oligo 5)

were used for the second round of 5' RACE PCR and the PCR products were analyzed as 10 above. A third round of 5' RACE PCR was carried out utilizing antisense primers:

5'-TGGAGCATGGTACGGGAATGCAGAAG-3' (SEQ.ID.NO.: 51; oligo 6) and

5'-GTGATGAGCAGGTCACTGAGCGCCAAG-3' (SEQ.ID.NO.: 52; oligo 7).

The sequence of the 5' RACE PCR products revealed the presence of the initiation codon ATG, and further round of 5' RACE PCR did not generate any more 5' sequence. The 15 completed 5' sequence was confirmed by RT-PCR using sense primer

5'-GCAATGCAGGCCTAACATTAC-3' (SEQ.ID.NO.: 53; oligo 8)

and oligo 4 as primers and sequence analysis of the 650 bp PCR product generated from human brain and heart cDNA templates (Clontech, Cat# 7404-1). The completed 3' sequence was confirmed by RT-PCR using oligo 2 and the following antisense primer:

20 5'-TTGGGTTACAATCTGAAGGGCA-3' (SEQ.ID.NO.: 54; oligo 9)

and sequence analysis of the 670 bp PCR product generated from human brain and heart cDNA templates. (Clontech, Cat# 7404-1).

d. hRUP5 (Seq. Id. Nos. 9 & 10)

The full length hRUP5 was cloned by RT-PCR using a sense primer upstream from

- 22 -

ATG, the initiation codon (SEQ.ID.NO.: 55), and an antisense primer containing TCA as the stop codon (SEQ.ID.NO.: 56), which had the following sequences:

5'-ACTCCGTGTCCAGCAGGACTCTG-3' (SEQ.ID.NO.:55)

5'-TGCCTGTTCCCTGGACCCTCACGTG-3' (SEQ.ID.NO.: 56)

5 and human peripheral leukocyte cDNA (Clontech) as a template. Advantage cDNA polymerase (Clontech) was used for the amplification in a 50ul reaction by the following cycle with step 2 through step 4 repeated 30 times: 94°C for 30 sec; 94° for 15 sec; 69° for 40 sec; 72°C for 3 min; and 72°C fro 6 min. A 1.4kb PCR fragment was isolated and cloned with the pCRII-TOPO™ vector (Invitrogen) and completely sequenced using the T7 DNA 10 Sequenase™ kit (Amsham). *See*, SEQ.ID.NO.: 9.

e. hRUP6 (Seq. Id. Nos. 11 & 12)

The full length hRUP6 was cloned by RT-PCR using primers:

5'-CAGGCCTTGGATTTAATGTCAGGGATGG-3' (SEQ.ID.NO.: 57) and

5'-GGAGAGTCAGCTCTGAAAGAATTCAAGG-3' (SEQ.ID.NO.: 58);

15 and human thymus Marathon-Ready™ cDNA (Clontech) as a template. Advantage cDNA polymerase (Clontech, according to manufacturer's instructions) was used for the amplification in a 50ul reaction by the following cycle: 94°C for 30sec; 94°C for 5 sec; 66°C for 40sec; 72°C for 2.5 sec and 72 °C for 7 min. Cycles 2 through 4 were repeated 30 times. A 1.3 Kb PCR fragment was isolated and cloned into the pCRII-TOPO™ vector (Invitrogen) 20 and completely sequenced (*see*, SEQ.ID.NO.: 11) using the ABI Big Dye Terminator™ kit (P.E. Biosystem).

f. hRUP7 (Seq. Id. Nos. 13 & 14)

The full length RUP7 was cloned by RT-PCR using primers:

5'-TGATGTGATGCCAGATACTAATAGCAC-3' (SEQ.ID.NO.: 59; sense) and

- 23 -

5'-CCTGATTCA~~TTAGGTGAGATTGAGAC~~-3' (SEQ.ID.NO.: 60; antisense) and human peripheral leukocyte cDNA (Clontech) as a template. Advantage™ cDNA polymerase (Clontech) was used for the amplification in a 50 ul reaction by the following cycle with step 2 to step 4 repeated 30 times: 94 °C for 2 minutes; 94 °C for 15 seconds; 60 °C 5 for 20 seconds; 72 °C for 2 minutes; 72 °C for 10 minutes. A 1.25 Kb PCR fragment was isolated and cloned into the pCRII-TOPO™ vector (Invitrogen) and completely sequenced using the ABI Big Dye Terminator™ kit (P.E. Biosystem). See, SEQ.ID.NO.: 13.

g. hARE-5 (Seq. Id. Nos. 5 & 6)

The full length hARE-5 was cloned by PCR using the hARE5 specific primers 10 5'-CAGCGCAGGGTGAAGCCTGAGAGC-3' SEQ.ID.NO.: 69 (sense, 5' of initiation codon ATG) and 5'-GGCACCTGCTGTGACCTGTGCAGG-3' SEQ.ID.NO.:70 (antisense, 3' of stop codon TGA) and human genomic DNA as template. TaqPlus Precision™ DNA polymerase (Stratagene) was used for the amplification by the following cycle with step 2 to step 4 repeated 35 times: 96°C, 2 minutes; 96°C, 20 seconds; 58°C, 30 seconds; 72°C, 2 minutes; and 72°C, 10 minutes 15 A 1.1 Kb PCR fragment of predicated size was isolated and cloned into the pCRII-TOPO™ vector (Invitrogen) and completely sequenced (SEQ.ID.NO.:5) using the T7 DNA Sequenase™ kit (Amsham).

h. hARE-4 (Seq. Id. Nos.: 3 & 4)

The full length hARE-4 was cloned by PCR using the hARE-4 specific primers 5'- 20 CTGGTGTGCTCCATGGCATCCC-3' SEQ.ID.NO.:67 (sense, 5' of initiation codon ATG) and 5'- GTAAGCCTCCCAGAACGAGAGG-3' SEQ.ID.NO.: 68 (antisense, 3' of stop codon TGA) and human genomic DNA as template. Taq DNA polymerase (Stratagene) and 5% DMSO was used for the amplification by the following cycle with step 2 to step 3 repeated 35 times:

- 24 -

94°C, 3 minutes; 94°C, 30 seconds; 59°C, 2 minutes; 72°C, 10 minutes

A 1.12 Kb PCR fragment of predicated size was isolated and cloned into the pCRII-TOPO™ vector (Invitrogen) and completely sequenced (SEQ.ID.NO.:3) using the T7 DNA Sequenase™ kit (Amsham).

5 i. **hARE-3 (Seq.Id.Nos.: 1 & 2)**

The full length hARE-3 was cloned by PCR using the hARE-3 specific primers 5'-gatcaagcttCCATCCTACTGAAACCATGGTC-3' SEQ.ID.NO.:65 (sense, lower case nucleotides represent Hind III overhang, **ATG** as initiation codon) and 5'-gatcagatctCAGTTCCAATATTCACACCACCGTC-3' SEQ.ID.NO.:66 (antisense, lower case 10 nucleotides represent Xba I overhang, **TCA** as stop codon) and human genomic DNA as template. TaqPlus Precision™ DNA polymerase (Stratagene) was used for the amplification by the following cycle with step 2 to step 4 repeated 35 times: 94°C, 3 minutes; 94°C, 1 minute; 55°C, 1 minute; 72°C, 2 minutes; 72°C, 10 minutes.

A 1.3 Kb PCR fragment of predicated size was isolated and digested with Hind III 15 and Xba I, cloned into the pRC/CMV2 vector (Invitrogen) at the Hind III and Xba I sites and completely sequenced (SEQ.ID.NO.:1) using the T7 DNA Sequenase™ kit (Amsham).

j. **hRUP3 (Seq. Id. Nos.:7 & 8)**

The full length hRUP3 was cloned by PCR using the hRUP3 specific primers 5'-GTCCTGCCACTTCGAGACATGG-3' SEQ.ID.NO.:71 (sense, **ATG** as initiation codon) and 5'-20 GAAACTTCTCTGCCCTTACCGTC-3' SEQ.ID.NO.:72 (antisense, 3' of stop codon TAA) and human genomic DNA as template. TaqPlus Precision™ DNA polymerase (Stratagene) was used for the amplification by the following cycle with step 2 to step 4 repeated 35 times: 94°C, 3 minutes; 94°C, 1 minute; 58°C, 1 minute; 72°C, 2 minutes; 72°C, 10 minutes

- 25 -

A 1.0 Kb PCR fragment of predicated size was isolated and cloned into the pCRII-TOPO™ vector (Invitrogen) and completely sequenced (SEQ.ID.NO.: 7) using the T7 DNA sequenase kit (Amsham).

Example 2
5 RECEPTOR EXPRESSION

Although a variety of cells are available to the art for the expression of proteins, it is most preferred that mammalian cells be utilized. The primary reason for this is predicated upon practicalities, *i.e.*, utilization of, *e.g.*, yeast cells for the expression of a GPCR, while possible, introduces into the protocol a non-mammalian cell which may not (indeed, in the 10 case of yeast, does not) include the receptor-coupling, genetic-mechanism and secretary pathways that have evolved for mammalian systems - thus, results obtained in non-mammalian cells, while of potential use, are not as preferred as that obtained from mammalian cells. Of the mammalian cells, COS-7, 293 and 293T cells are particularly preferred, although the specific mammalian cell utilized can be predicated upon the particular 15 needs of the artisan. The general procedure for expression of the disclosed GPCRs is as follows.

On day one, 1×10^7 293T cells per 150mm plate were plated out. On day two, two reaction tubes will be prepared (the proportions to follow for each tube are per plate): tube A will be prepared by mixing 20 μ g DNA (*e.g.*, pCMV vector; pCMV vector with receptor 20 cDNA, etc.) in 1.2ml serum free DMEM (Irvine Scientific, Irvine, CA); tube B will be prepared by mixing 120 μ l lipofectamine (Gibco BRL) in 1.2ml serum free DMEM. Tubes A and B are admixed by inversions (several times), followed by incubation at room temperature for 30-45min. The admixture can be referred to as the "transfection mixture". Plated 293T cells are washed with 1XPBS, followed by addition of 10ml serum free DMEM.

- 26 -

2.4ml of the transfection mixture will then be added to the cells, followed by incubation for 4hrs at 37°C/5% CO₂. The transfection mixture was then be removed by aspiration, followed by the addition of 25ml of DMEM/10% Fetal Bovine Serum. Cells will then be incubated at 37°C/5% CO₂. After 72hr incubation, cells can then be harvested and utilized for analysis.

5 Example 3

TISSUE DISTRIBUTION OF THE DISCLOSED HUMAN GPCRS

Several approaches can be used for determination of the tissue distribution of the GPCRs disclosed herein.

1. Dot-Blot Analysis

10 Using a commercially available human-tissue dot-blot format, endogenous orphan GPCRs were probed for a determination of the areas where such receptors are localized. cDNA fragments from the GPCRs of Example 1 (radiolabelled) were (or can be) used as the probe: radiolabeled probe was (or can be) generated using the complete receptor cDNA (excised from the vector) using a Prime-It II™ Random Primer Labeling Kit (Stratagene, 15 #300385), according to manufacturer's instructions. A human RNA Master Blot™ (Clontech, #7770-1) was hybridized with the endogenous human GPCR radiolabeled probe and washed under stringent conditions according manufacturer's instructions. The blot was exposed to Kodak BioMax™ Autoradiography film overnight at -80°C. Results are summarized for several receptors in Table B and C (see Figures 1A and 1B for a grid 20 identifying the various tissues and their locations, respectively). Exemplary dot-blots are provided in Figure 2A and 2B for results derived using hCHN3 and hCHN8, respectively.

TABLE B

ORPHAN GPCR

Tissue Distribution (highest levels, relative to other tissues in the dot-blot)

- 27 -

	hGPCR27	Fetal brain, Putamen, Pituitary gland, Caudate nucleus
	hARE-1	Spleen, Peripheral leukocytes, Fetal spleen
	hPPR1	Pituitary gland, Heart, salivary gland, Small intestine, Testis
	hRUP3	Pancreas
5	hCHN3	Fetal brain, Putamen, Occipital cortex
	hCHN9	Pancreas, Small intestine, Liver
	hCHN10	Kidney, Thryoid

TABLE C

	ORPHAN GPCR	Tissue Distribution (highest levels, relative to other tissues in the dot-blot)
10	hARE-3	Cerebellum left, Cerebellum right, Testis, Accumbens
	hGPCR3	Corpus collusum, Caudate nucleus, Liver, Heart, Inter-Ventricular Septum
	hARE-2	Cerebellum left, Cerebellum right, Substantia
	hCHN8	Cerebellum left, Cerebellum right, Kidney, Lung

2. RT-PCR

15 a. hRUP3

To ascertain the tissue distribution of hRUP3 mRNA, RT-PCR was performed using hRUP3-specific primers and human multiple tissue cDNA panels (MTC, Clontech) as templates. Taq DNA polymerase (Stratagene) was utilized for the PCR reaction, using the following reaction cycles in a 40ul reaction: 94 °C for 2 min; 94 °C for 15 sec; 55 °C for 30 sec; 72 °C for 1 min; 72 ° C, for 10 min. Primers were as follows:

5'-GACAGGTACCTGCCATCAAG-3' (SEQ.ID.NO.: 61; sense)

5'-CTGCACAATGCCAGTGATAAGG-3' (SEQ.ID.NO.: 62; antisense).

20ul of the reaction was loaded onto a 1% agarose gel; results are set forth in Figure 3.

As is supported by the data of Figure 3, of the 16 human tissues in the cDNA panel utilized (brain, colon, heart, kidney, lung, ovary, pancreas, placenta, prostate, skeleton, small intestine, spleen, testis, thymus leukocyte, and liver) a single hRUP3 band is evident only from the pancreas. Additional comparative analysis of the protein sequence of hRUP3 with 5 other GPCRs suggest that hRUP3 is related to GPCRs having small molecule endogenous ligand such that it is predicted that the endogenous ligand for hRUP3 is a small molecule.

b. hRUP4

RT-PCR was performed using hRUP4 oligo's 8 and 4 as primers and the human multiple tissue cDNA panels (MTC, Clontech) as templates. Taq DNA polymerase 10 (Stratagene) was used for the amplification in a 40ul reaction by the following cycles: 94°C for 30 seconds, 94°C for 10 seconds, 55°C for 30 seconds, 72°C for 2 minutes, and 72°C for 5 minutes with cycles 2 through 4 repeated 30 times.

20 μ l of the reaction were loaded on a 1% agarose gel to analyze the RT-PCR products, and hRUP4 mRNA was found expressed in many human tissues, with the strongest 15 expression in heart and kidney. (see, Figure 4). To confirm the authenticity of the PCR fragments, a 300 bp fragment derived from the 5' end of hRUP4 was used as a probe for the Southern Blot analysis. The probe was labeled with 32 P-dCTP using the Prime-It IITM Random Primer Labeling Kit (Stratagene) and purified using the ProbeQuantTM G-50 micro columns (Amersham). Hybridization was done overnight at 42° C following a 12 hr pre-20 hybridization. The blot was finally washed at 65°C with 0.1 x SSC. The Southern blot did confirm the PCR fragments as hRUP4.

c. hRUP5

- 29 -

RT-PCR was performed using the following hRUP5 specific primers:

5'-CTGACTTCTTGTCTGGCAGCAGCGG-3' (SEQ.ID.NO.: 63; sense)

5'-AGACCAGCCAGGGCACGCTGAAGAGTG-3' (SEQ.ID.NO.: 64; antisense)

and the human multiple tissue cDNA panels (MTC, Clontech) as templates. Taq DNA polymerase (Stratagene) was used for the amplification in a 40ul reaction by the following cycles: 94°C for 30 sec, 94°C for 10 sec, 62°C for 1.5 min, 72°C for 5 min, and with cycles 2 through 3 repeated 30 times. 20 µl of the reaction were loaded on a 1.5% agarose gel to analyze the RT-PCR products, and hRUP5 mRNA was found expressed only in the peripheral blood leukocytes (*data not shown*).

10 d. hRUP6

RT-PCR was applied to confirm the expression and to determine the tissue distribution of hRUP6. Oligonucleotides used, based on an alignment of AC005871 and GPR66 segments, had the following sequences:

5'-CCAACACCAGCATCCATGGCATCAAG-3' (SEQ.ID.NO.: 73; sense),

15 5'-GGAGAGTCAGCTCTGAAAGAATTCAAGG-3' (SEQ.ID.NO.: 74; antisense)

and the human multiple tissue cDNA panels (MTC, Clontech) were used as templates.

PCR was performed using TaqPlus Precision™ polymerase (Stratagene; manufacturing instructions will be followed) in a 40ul reaction by the following cycles: 94°C for 30 sec; 94°C 5 sec; 66°C for 40 sec, 72°C for 2.5 min, and 72°C for 7 min. Cycles 2 through 4 were repeated 30 times.

20 µl of the reaction were loaded on a 1.2% agarose gel to analyze the RT-PCR products, and a specific 760bp DNA fragment representing hRUP6 was expressed predominantly in the thymus and with less expression in the heart, kidney, lung, prostate and small intestine and testis. (see, Figure 5).

- 30 -

It is intended that each of the patents, applications, and printed publications mentioned in this patent document be hereby incorporated by reference in their entirety.

As those skilled in the art will appreciate, numerous changes and modifications may be made to the preferred embodiments of the invention without departing from the spirit of the invention. It is intended that all such variations fall within the scope of the invention and the claims that follow.

Although a variety of Vectors are available to those in the art, for purposes of utilization for both endogenous and non-endogenous human GPCRs, it is most preferred that the Vector utilized be pCMV. This vector was deposited with the American Type 10 Culture Collection (ATCC) on October 13, 1998 (10801 University Blvd., Manassas, VA 20110-2209 USA) under the provisions of the Budapest Treaty for the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure. The DNA was tested by the ATCC and determined to be. The ATCC has assigned the following deposit number to pCMV: ATCC #203351.

CLAIMS

What is claimed is:

1. A cDNA encoding a human G protein-coupled receptor comprising SEQ.ID.NO.: 1.
- 5 2. A human G protein-coupled receptor encoded by the cDNA of SEQ.ID.NO.: 1 comprising SEQ.ID.NO.: 2.
3. A Plasmid comprising a Vector and the cDNA of SEQ.ID.NO.:1.
4. A Host Cell comprising the Plasmid of claim 3.
5. A cDNA encoding a human G protein-coupled receptor comprising 10 SEQ.ID.NO.: 3.
6. A human G protein-coupled receptor encoded by the cDNA of SEQ.ID.NO.: 3 comprising SEQ.ID.NO.: 4.
7. A Plasmid comprising a Vector and the cDNA of SEQ.ID.NO.:3.
8. A Host Cell comprising the Plasmid of claim 7.
- 15 9. A cDNA encoding a human G protein-coupled receptor comprising SEQ.ID.NO.: 5.
10. A human G protein-coupled receptor encoded by the cDNA of SEQ.ID.NO.: 5 comprising SEQ.ID.NO.: 6.
11. A Plasmid comprising a Vector and the cDNA of SEQ.ID.NO.:5.
- 20 12. A Host Cell comprising the Plasmid of claim 11.
13. A cDNA encoding a human G protein-coupled receptor comprising SEQ.ID.NO.: 7.

- 32 -

14. A human G protein-coupled receptor encoded by the cDNA of SEQ.ID.NO.: 7 comprising SEQ.ID.NO.: 8.
15. A Plasmid comprising a Vector and the cDNA of SEQ.ID.NO.:7.
16. A Host Cell comprising the Plasmid of claim 15.
- 5 17. A cDNA encoding a human G protein-coupled receptor comprising SEQ.ID.NO.: 9.
18. A human G protein-coupled receptor encoded by the cDNA of SEQ.ID.NO.: 9 comprising SEQ.ID.NO.: 10.
19. A Plasmid comprising a Vector and the cDNA of SEQ.ID.NO.:9.
- 10 20. A Host Cell comprising the Plasmid of claim 19.
21. A cDNA encoding a human G protein-coupled receptor comprising SEQ.ID.NO.: 11.
22. A human G protein-coupled receptor encoded by the cDNA of SEQ.ID.NO.: 11 comprising SEQ.ID.NO.:12.
- 15 23. A Plasmid comprising a Vector and the cDNA of SEQ.ID.NO.:11.
24. A Host Cell comprising the Plasmid of claim 23.
25. A cDNA encoding a human G protein-coupled receptor comprising SEQ.ID.NO.: 13.
26. A human G protein-coupled receptor encoded by the cDNA of 20 SEQ.ID.NO.: 13 comprising SEQ.ID.NO.: 14.
27. A Plasmid comprising a Vector and the cDNA of SEQ.ID.NO.:13.
28. A Host Cell comprising the Plasmid of claim 27.
29. A cDNA encoding a human G protein-coupled receptor comprising

SEQ.ID.NO.: 15.

30. A human G protein-coupled receptor encoded by the cDNA of SEQ.ID.NO.: 15 comprising SEQ.ID.NO.: 16.
31. A Plasmid comprising a Vector and the cDNA of SEQ.ID.NO.:15.
- 5 32. A Host Cell comprising the Plasmid of claim 31.
33. A cDNA encoding a human G protein-coupled receptor comprising SEQ.ID.NO.: 17.
34. A human G protein-coupled receptor encoded by the cDNA of SEQ.ID.NO.: 17 comprising SEQ.ID.NO.: 18.
- 10 35. A Plasmid comprising a Vector and the cDNA of SEQ.ID.NO.:17.
36. A Host Cell comprising the Plasmid of claim 35.
37. A cDNA encoding a human G protein-coupled receptor comprising SEQ.ID.NO.: 19.
38. A human G protein-coupled receptor encoded by the cDNA of 15 SEQ.ID.NO.: 19 comprising SEQ.ID.NO.: 20.
39. A Plasmid comprising a Vector and the cDNA of SEQ.ID.NO.:19.
40. A Host Cell comprising the Plasmid of claim 39.
41. A cDNA encoding a human G protein-coupled receptor comprising SEQ.ID.NO.: 21.
- 20 42. A human G protein-coupled receptor encoded by the cDNA of SEQ.ID.NO.: 21 comprising SEQ.ID.NO.: 22.
43. A Plasmid comprising a Vector and the cDNA of SEQ.ID.NO.:21.
44. A Host Cell comprising the Plasmid of claim 43.

- 34 -

45. A cDNA encoding a human G protein-coupled receptor comprising SEQ.ID.NO.: 23.
46. A human G protein-coupled receptor encoded by the cDNA of SEQ.ID.NO.: 23 comprising SEQ.ID.NO.: 24.
- 5 47. A Plasmid comprising a Vector and the cDNA of SEQ.ID.NO.: 23.
48. A Host Cell comprising the Plasmid of claim 47.
49. A cDNA encoding a human G protein-coupled receptor comprising SEQ.ID.NO.: 25.
50. A human G protein-coupled receptor encoded by the cDNA of 10 SEQ.ID.NO.: 25 comprising SEQ.ID.NO.: 26.
51. A Plasmid comprising a Vector and the cDNA of SEQ.ID.NO.:25.
52. A Host Cell comprising the Plasmid of claim 51.
53. A cDNA encoding a human G protein-coupled receptor comprising SEQ.ID.NO.: 27.
- 15 54. A human G protein-coupled receptor encoded by the cDNA of SEQ.ID.NO.: 27 comprising SEQ.ID.NO.: 28.
55. A Plasmid comprising a Vector and the cDNA of SEQ.ID.NO.:27.
56. A Host Cell comprising the Plasmid of claim 55.
57. A cDNA encoding a human G protein-coupled receptor comprising 20 SEQ.ID.NO.: 29.
58. A human G protein-coupled receptor encoded by the cDNA of SEQ.ID.NO.: 29 comprising SEQ.ID.NO.: 30.
59. A Plasmid comprising a Vector and the cDNA of SEQ.ID.NO.:29.

- 35 -

60. A Host Cell comprising the Plasmid of claim 59.
61. A cDNA encoding a human G protein-coupled receptor comprising SEQ.ID.NO.: 31.
62. A human G protein-coupled receptor encoded by the cDNA of 5 SEQ.ID.NO.: 31 comprising SEQ.ID.NO.: 32.
63. A Plasmid comprising a Vector and the cDNA of SEQ.ID.NO.:31.
64. A Host Cell comprising the Plasmid of claim 63.
65. A cDNA encoding a human G protein-coupled receptor comprising SEQ.ID.NO.: 33.

10 66. A human G protein-coupled receptor encoded by the cDNA of SEQ.ID.NO.: 33 comprising SEQ.ID.NO.: 34.

67. A Plasmid comprising a Vector and the cDNA of SEQ.ID.NO.:33.
68. A Host Cell comprising the Plasmid of claim 67.
69. A cDNA encoding a human G protein-coupled receptor comprising 15 SEQ.ID.NO.: 35.

70. A human G protein-coupled receptor encoded by the cDNA of SEQ.ID.NO.: 35 comprising SEQ.ID.NO.: 36.

71. A Plasmid comprising a Vector and the cDNA of SEQ.ID.NO.:35.
72. A Host Cell comprising the Plasmid of claim 71.

20 73. A cDNA encoding a human G protein-coupled receptor comprising SEQ.ID.NO.: 37.

74. A human G protein-coupled receptor encoded by the cDNA of SEQ.ID.NO.: 37 comprising SEQ.ID.NO.: 38.

- 36 -

75. A Plasmid comprising a Vector and the cDNA of SEQ.ID.NO.:37.

76. A Host Cell comprising the Plasmid of claim 75.

1/4

	1	2	3	4	5	6	7	8
A		Amygdala	Caudate	Cerebellum	Cerebral Cortex	Frontal Cortex	Hippocampus	Medulla Oblongata
B	Occipital Cortex	Putamen	Substantia Nigra	Temporal Cortex	Thalamus	Accumbens	Spinal Cord	
C	Heart	Aorta	Skeletal Muscle	Colon	Bladder	Uterus	Prostate	Stomach
D	Testis	Ovary	Pancreas	Pituitary	Adrenal Gland	Thyroid	Salivary Gland	Mammary Gland
E	Kidney	Liver	Small Intestine	Spleen	Thymus	Peripheral Leukocyte	Lymph Node	Bone Marrow
F	Appendix	Lung	Trachea	Placenta				
G	Fetal Brain	Fetal Heart	Fetal Kidney	Fetal Liver	Fetal Spleen	Fetal Thymus	Fetal Lung	
H								

FIG. 1A

2 / 4

	1	2	3	4	5	6	7	8	9	10	11	12
A	Cerebellum Left	Substantia Nigra	Heart	Esophagus	Colon Transverse	Kidney	Lung	Liver	Leukemia	Fetal Brain		
B	Cerebral Cortex	Cerebellum Right	Accumbens	Aorta	Stomach	Colon Descending	Skeletal Muscle	Placenta	Pancreas	HeLa S3	Fetal Heart	
C	Frontal Cortex	Corpus Callosum	Thalamus	Atrium Left	Duodenum	Rectum	Spleen	Bladder	Adrenal Gland	Leukemia	Fetal Kidney	
D	Parietal Lobe	Amygdala	Pituitary Gland	Atrium Right	Jejunum	Thymus	Uterus	Thyroid	MOLT-4	Leukemia	Fetal Liver	
E	Occipital Cortex	Claudete Nucleus	Spinal Cord	Ventricle Left	Ileum	Peripheral Leukocyte	Prostate	Salivary Gland	Burkitt's Lymphoma	Burkitt's Lymphoma	Fetal Spleen	
F	Temporal Cortex	Hippocampus		Ventricle Right	Ileocecum	Lymph Node	Testis	Mammary Gland	Raji	Burkitt's Lymphoma	Fetal Thymus	
G	Paracentral Gyrus of Cerebral Cortex	Medulla Oblongata		Inter Ventricular Septum	Appendix	Bone Marrow	Ovary		Daudi	Colorectal Adenocarcinoma	Fetal Lung	
H	Pons	Putamen		Apex of the Heart	Colon Ascending		Trachea			Lung	Carcinoma	A549

3/4

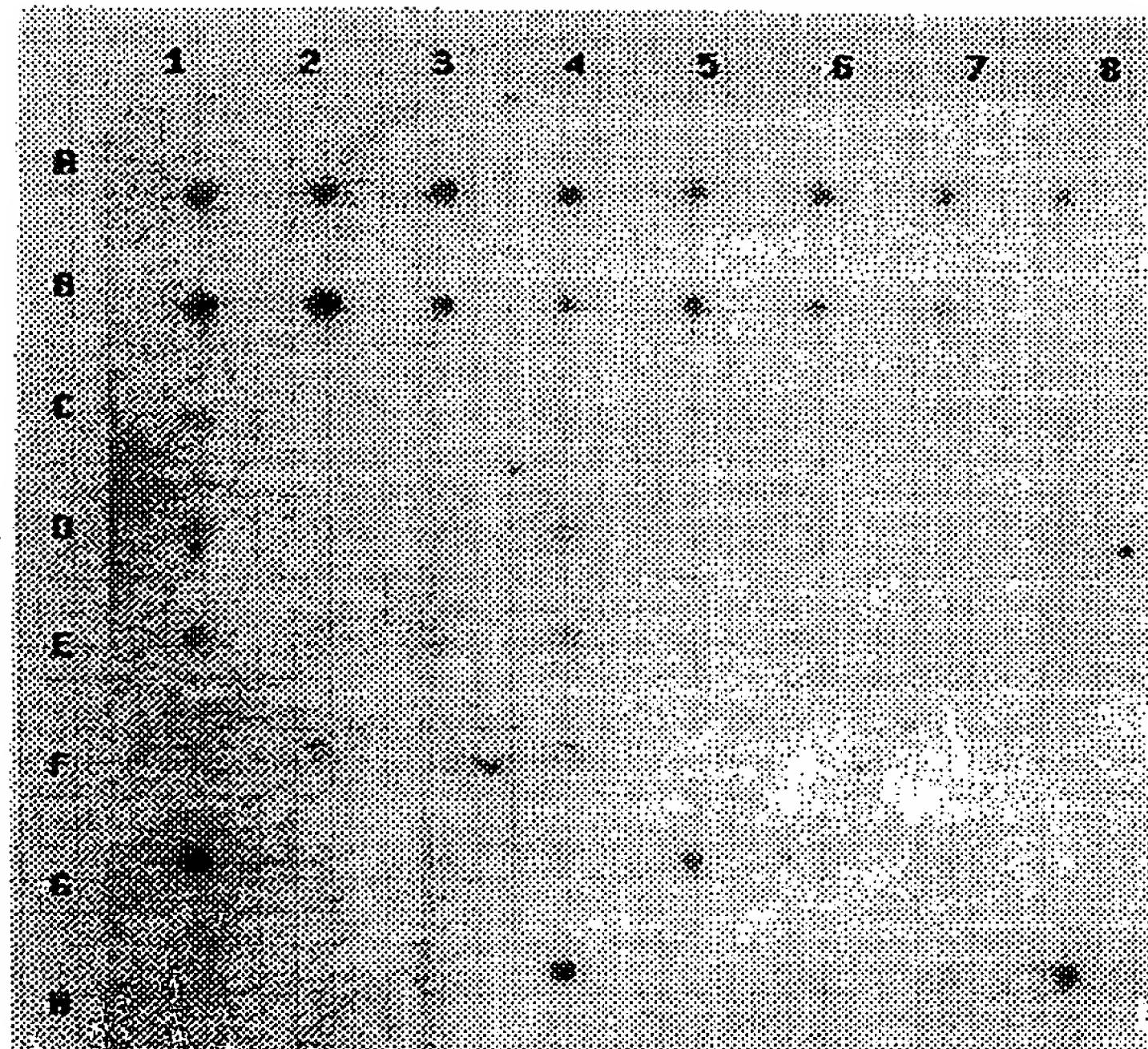


FIG. 2A

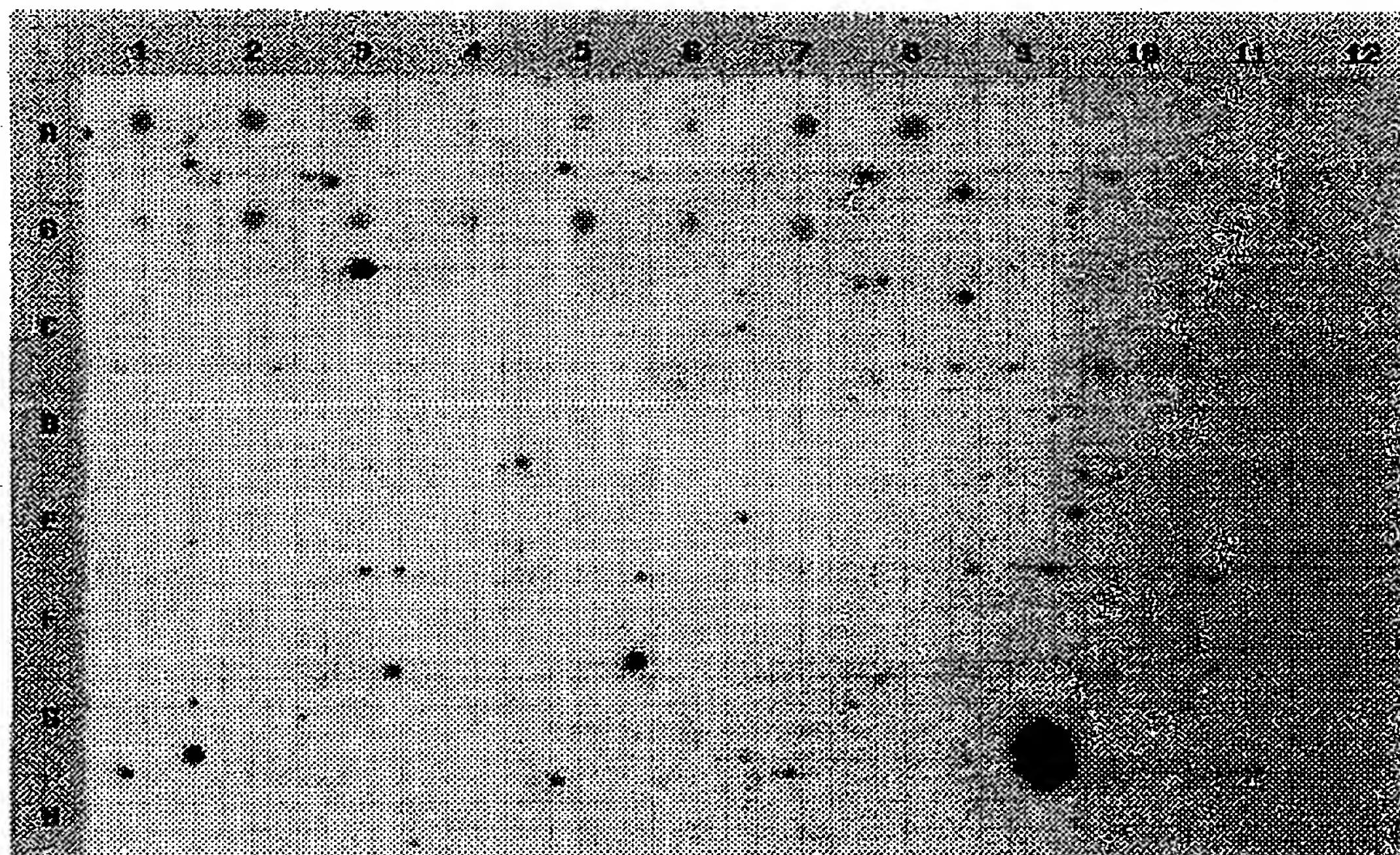


FIG. 2B

FIG. 3



FIG. 4

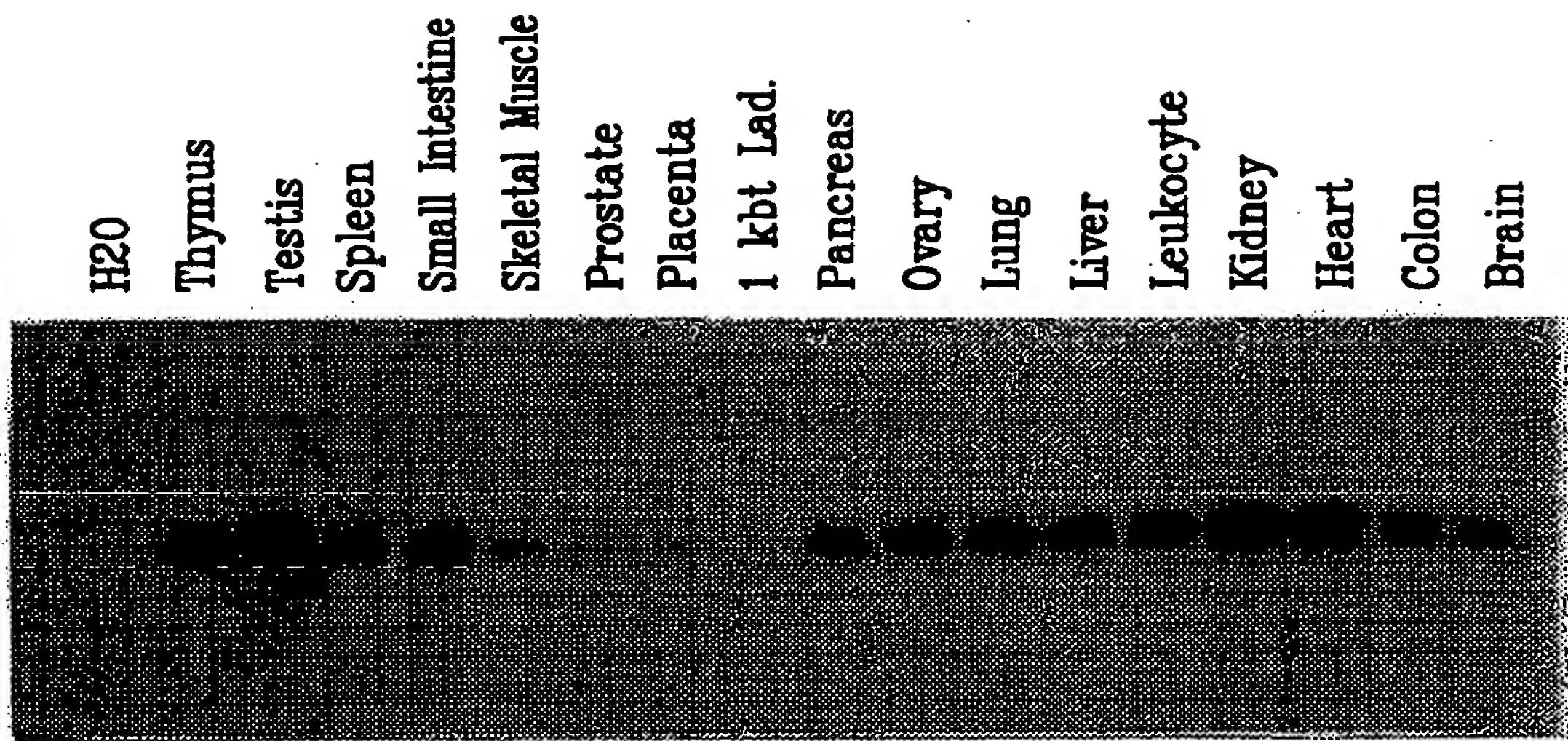
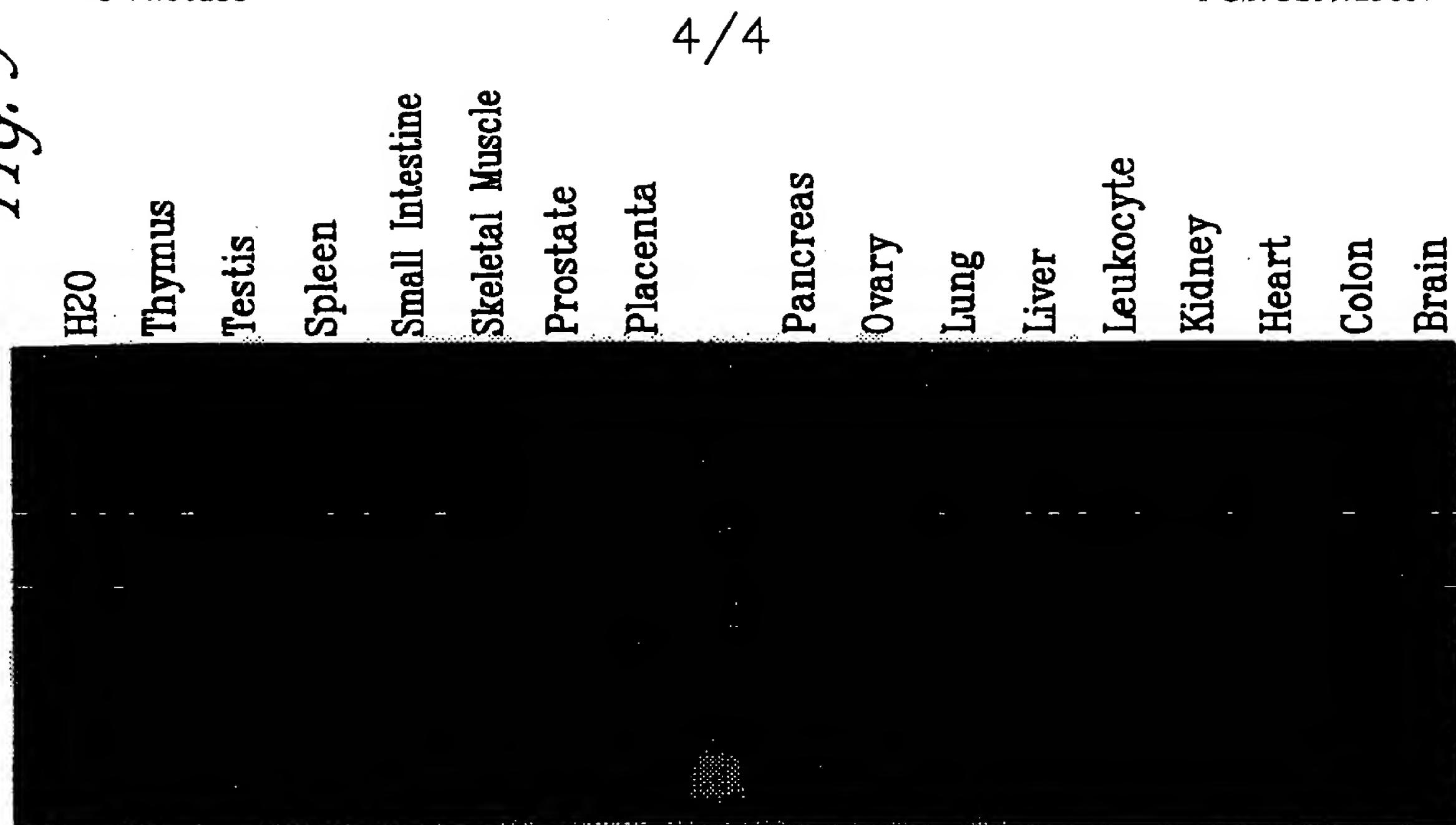


FIG. 5



- 1 -

SEQUENCE LISTING

(1) GENERAL INFORMATION:

5 (i) APPLICANT: Chen, Ruoping
Dang, Huong T.
Liaw, Chen W.
Lin, I-Lin

(ii) TITLE OF INVENTION: Human Orphan G Protein-Coupled Receptors

10 (iii) NUMBER OF SEQUENCES: 74

15 (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Arena Pharmaceuticals, Inc.
(B) STREET: 6166 Nancy Ridge Drive
(C) CITY: San Diego
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 92121

20 (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

25 (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US
(B) FILING DATE:
(C) CLASSIFICATION:

30 (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Burgoon, Richard P.
(B) REGISTRATION NUMBER: 34,787

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (858) 453-7200
(B) TELEFAX: (858) 453-7210

(2) INFORMATION FOR SEQ ID NO:1:

35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1260 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

40 ATGGTCTTCT CGGCAGTGTT GACTGCGTTC CATACCGGGA CATCCAACAC AACATTTGTC 60

- 2 -

GTGTATGAAA ACACCTACAT GAATATTACA CTCCCTCCAC CATTCCAGCA TCCTGACCTC 120
AGTCCATTGC TTAGATATAG TTTGAAACC ATGGCTCCA CTGGTTGAG TTCCTTGACC 180
GTGAATAGTA CAGCTGTGCC CACAACACCA GCAGCATTAA AGAGCCTAAA CTTGCCTCTT 240
CAGATCACCC TTTCTGCTAT AATGATATTC ATTCTGTTG TGTCTTTCT TGGGAACCTG 300
5 GTTGTGTTGCC TCATGGTTA CCAAAAAGCT GCCATGAGGT CTGCAATTAA CATCCTCCTT 360
GCCAGCCTAG CTTTGAGA CATGTTGCTT GCAGTGCTGA ACATGCCCTT TGCCCTGGTA 420
ACTATTCTTA CTACCCGATG GATTTTGGG AAATTCTTCT GTAGGGTATC TGCTATGTTT 480
TTCTGGTTAT TTGTGATAGA AGGAGTAGCC ATCCTGCTCA TCATTAGCAT AGATAGGGTC 540
CTTATTATAG TCCAGAGGCA GGATAAGCTA AACCCATATA GAGCTAAGGT TCTGATTGCA 600
10 GTTTCTTGGG CAACTTCCTT TTGTGTAGCT TTTCTTTAG CCGTAGGAAA CCCCACCTG 660
CAGATACCTT CCCGAGCTCC CCAGTGTGTG TTTGGGTACA CAACCAATCC AGGCTACCAAG 720
GCTTATGTGA TTTGATTTC TCTCATTCT TTCTTCATAC CCTTCCTGGT AATACTGTAC 780
TCATTTATGG GCATACTCAA CACCCCTCGG CACAATGCCT TGAGGATCCA TAGCTACCC 840
GAAGGTATAT GCCTCAGCCA GGCCAGCAAA CTGGGTCTCA TGAGTCTGCA GAGACCTTC 900
15 CAGATGAGCA TTGACATGGG CTTTAAACA CGTGCCTCA CCACTATTT GATTCTCTT 960
GCTGTCTCA TTGTCTGCTG GGCCCCATTC ACCACTTACA GCCTTGTGGC
AACATTCACTG 1020
AAGCACTTT ACTATCAGCA CAACTTTTT GAGATTAGCA CCTGGCTACT GTGGCTCTGC 1080
TACCTCAAGT CTGCATTGAA TCCGCTGATC TACTACTGGA GGATTAAGAA ATTCCATGAT 1140
20 GCTTGCTGG ACATGATGCC TAAGTCCTTC AAGTTTTGC CGCAGCTCCC TGGTCACACA 1200
AAGCGACGGA TACGTCTAG TGCTGTCTAT GTGTGTGGGG AACATCGGAC GGTGGTGTGA 1260

(3) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 419 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

30 Met Val Phe Ser Ala Val Leu Thr Ala Phe His Thr Gly Thr Ser Asn
1 5 10 15

- 12 -

CTCTTCAGCG TGCCCTGGCT GGTCTTCCCC GAGGCTGCCG TCTGGTGGTA CGACCTGGTC 540
 ATCTGCCTGG ACTTCTGGGA CAGCGAGGAG CTGTCGCTGA GGATGCTGGA GGTCTGGGG 600
 GGCTTCCTGC CTTTCCTCCT GCTGCTCGTC TGCCACGTGC TCACCCAGGC CACAGCCTGT 660
 CGCACCTGCC ACCGCCAACA GCAGCCCGCA GCCTGCCGGG GCTTCGCCCG TGTGGCCAGG 720
 5 ACCATTCTGT CAGCCTATGT GGTCTGAGG CTGCCCTACC AGCTGGCCA GCTGCTCTAC 780
 CTGGCCTTCC TGTGGGACGT CTACTCTGGC TACCTGCTCT GGGAGGCCCT GGTCTACTCC 840
 GACTACCTGA TCCTACTCAA CAGCTGCCTC AGCCCCTTCC TCTGCCTCAT GGCCAGTGCC 900
 GACCTCCGGA CCCTGCTGCG CTCCGTGCTC TCGTCCTTCG CGGCAGCTCT CTGCGAGGAG 960
 CGGCCGGGCA GCTTCACGCC CACTGAGCCA CAGACCCAGC TAGATTCTGA GGGTCCA 1020
 10 CTGCCAGAGC CGATGGCAGA GGCCCAGTCA CAGATGGATC CTGTGGCCA GCCTCAGGTG 1080
 AACCCCCACAC TCCAGCCACG ATCGGATCCC ACAGCTCAGC CACAGCTGAA CCCTACGGCC 1140
 CAGCCACAGT CGGATCCCAC AGCCCAGCCA CAGCTGAACC TCATGGCCA GCCACAGTCA 1200
 GATTCTGTGG CCCAGCCACA GGCAGACACT AACGTCCAGA CCCCTGCACC TGCTGCCAGT 1260
 TCTGTGCCCA GTCCCTGTGA TGAAGCTTCC CCAACCCCAT CCTCGCATCC TACCCCAGGG 1320
 15 GCCCTTGAGG ACCCAGCCAC ACCTCCTGCC TCTGAAGGAG AAAGCCCCAG CAGCACCCCG 1380
 CCAGAGGCGG CCCCAGGCCAG TGA

1413

(11) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 468 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

25 Met Asp Thr Thr Met Glu Ala Asp Leu Gly Ala Thr Gly His Arg Pro
 1 5 10 15

Arg Thr Glu Leu Asp Asp Glu Asp Ser Tyr Pro Gln Gly Gly Trp Asp
 20 25 30

30 Thr Val Phe Leu Val Ala Leu Leu Leu Gly Leu Pro Ala Asn Gly
 35 40 45

Leu Met Ala Trp Leu Ala Gly Ser Gln Ala Arg His Gly Ala Gly Thr

- 13 -

	50	55	60
	Arg Leu Ala Leu Leu Leu Leu Ser	Leu Ala Leu Ser Asp Phe Leu Phe	
65	70	75	80
5	Leu Ala Ala Ala Ala Phe Gln Ile Leu Glu Ile Arg His Gly Gly His		
	85	90	95
	Trp Pro Leu Gly Thr Ala Ala Cys Arg Phe Tyr Tyr Phe Leu Trp Gly		
	100	105	110
10	Val Ser Tyr Ser Ser Gly Leu Phe Leu Leu Ala Ala Leu Ser Leu Asp		
	115	120	125
	Arg Cys Leu Leu Ala Leu Cys Pro His Trp Tyr Pro Gly His Arg Pro		
	130	135	140
	Val Arg Leu Pro Leu Trp Val Cys Ala Gly Val Trp Val Leu Ala Thr		
	145	150	155
	160		
15	Leu Phe Ser Val Pro Trp Leu Val Phe Pro Glu Ala Ala Val Trp Trp		
	165	170	175
	Tyr Asp Leu Val Ile Cys Leu Asp Phe Trp Asp Ser Glu Glu Leu Ser		
	180	185	190
	Leu Arg Met Leu Glu Val Leu Gly Gly Phe Leu Pro Phe Leu Leu Leu		
	195	200	205
20	Leu Val Cys His Val Leu Thr Gln Ala Thr Arg Thr Cys His Arg Gln		
	210	215	220
	Gln Gln Pro Ala Ala Cys Arg Gly Phe Ala Arg Val Ala Arg Thr Ile		
	225	230	235
	240		
25	Leu Ser Ala Tyr Val Val Leu Arg Leu Pro Tyr Gln Leu Ala Gln Leu		
	245	250	255
	Leu Tyr Leu Ala Phe Leu Trp Asp Val Tyr Ser Gly Tyr Leu Leu Trp		
	260	265	270
	Glu Ala Leu Val Tyr Ser Asp Tyr Leu Ile Leu Leu Asn Ser Cys Leu		
	275	280	285
30	Ser Pro Phe Leu Cys Leu Met Ala Ser Ala Asp Leu Arg Thr Leu Leu		
	290	295	300
	Arg Ser Val Leu Ser Ser Phe Ala Ala Ala Leu Cys Glu Glu Arg Pro		
	305	310	315
	320		
35	Gly Ser Phe Thr Pro Thr Glu Pro Gln Thr Gln Leu Asp Ser Glu Gly		
	325	330	335
	Pro Thr Leu Pro Glu Pro Met Ala Glu Ala Gln Ser Gln Met Asp Pro		
	340	345	350

- 14 -

	Val	Ala	Gln	Pro	Gln	Val	Asn	Pro	Thr	Leu	Gln	Pro	Arg	Ser	Asp	Pro
																355
																360
																365
	Thr	Ala	Gln	Pro	Gln	Leu	Asn	Pro	Thr	Ala	Gln	Pro	Gln	Ser	Asp	Pro
																370
																375
																380
5	Thr	Ala	Gln	Pro	Gln	Leu	Asn	Leu	Met	Ala	Gln	Pro	Gln	Ser	Asp	Ser
																385
																390
																395
																400
	Val	Ala	Gln	Pro	Gln	Ala	Asp	Thr	Asn	Val	Gln	Thr	Pro	Ala	Pro	Ala
																405
																410
																415
10	Ala	Ser	Ser	Val	Pro	Ser	Pro	Cys	Asp	Glu	Ala	Ser	Pro	Thr	Pro	Ser
																420
																425
																430
	Ser	His	Pro	Thr	Pro	Gly	Ala	Leu	Glu	Asp	Pro	Ala	Thr	Pro	Pro	Ala
																435
																440
																445
	Ser	Glu	Gly	Glu	Ser	Pro	Ser	Ser	Thr	Pro	Pro	Glu	Ala	Ala	Pro	Gly
																450
																455
																460
15	Ala	Gly	Pro	Thr												
																465

(12) INFORMATION FOR SEQ ID NO:11:

	(i) SEQUENCE CHARACTERISTICS:
20	(A) LENGTH: 1248 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: single
	(D) TOPOLOGY: linear
	(ii) MOLECULE TYPE: DNA (genomic)
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:
25	ATGTCAGGGA TGGAAAAACT TCAGAATGCT TCCTGGATCT ACCAGCAGAA ACTAGAAGAT 60
	CCATTCCAGA AACACCTGAA CAGCACCGAG GAGTATCTGG CCTTCCTCTG CGGACCTCGG 120
	CGCAGCCACT TCTTCCTCCC CGTGTCTGTG GTGTATGTGC CAATTTTGT GGTGGGGGTC 180
	ATTGGCAATG TCCTGGTGTG CCTGGTGATT CTGCAGCACC AGGCTATGAA GACGCCACC 240
	AACTACTACC TCTTCAGCCT GGCGGTCTCT GACCTCCTGG TCCTGCTCCT TGGAAATGCC 300
30	CTGGAGGTCT ATGAGATGTG GCGCAACTAC CCTTTCTTGT TCAGGGCCCGT GGGCTGCTAC 360
	TTCAAGACGG CCCTCTTGA GACCGTGTGC TTCCGCTCCA TCCTCAGCAT CACCACCGTC 420
	AGCGTGGAGC GCTACGTGGC CATCCTACAC CCGTTCCGCG CCAAACGTCA GAGCACCCGG 480
	CGCCGGGCC CGCAGGATCCT CGGCATCGTC TGGGGCTTCT CCGTGCTCTT CTCCCTGCC 540

- 15 -

AACACCAGCA TCCATGGCAT CAAGTTCCAC TACTTCCCCA ATGGGTCCCT GGTCCCAGGT 600
 TCGGCCACCT GTACGGTCAT CAAGCCCAGT TGGATCTACA ATTCATCAT CCAGGTCAACC 660
 TCCTTCCTAT TCTACCTCCT CCCCATGACT GTCATCAGTG TCCTCTACTA CCTCATGGCA 720
 CTCAGACTAA AGAAAGACAA ATCTCTTGAG GCAGATGAAG GGAATGCAAA TATTCAAAGA 780
 5 CCCTGCAGAA AATCAGTCAA CAAGATGCTG TTTGTCTTGG TCTTAGTGT TGCTATCTGT 840
 TGGGCCCGT TCCACATTGA CCGACTCTTC TTCAGCTTG TGGAGGAGTG GAGTGAATCC 900
 CTGGCTGCTG TGTTCAACCT CGTCCATGTG GTGTCAGGTG TCTTCTTCTA CCTGAGCTCA 960
 GCTGTCAACC CCATTATCTA TAACCTACTG TCTCGCCGCT TCCAGGCAGC ATTCCAGAAT1020
 GTGATCTCTT CTTTCCACAA ACAGTGGCAC TCCCAGCATG ACCCACAGTT GCCACCTGCC1080
 10 CAGCGGAACA TCTTCCTGAC AGAATGCCAC TTTGTGGAGC TGACCGAAGA TATAGGTCCC1140
 CAATTCCCAT GTCAGTCATC CATGCACAAAC TCTCACCTCC CAACAGCCCT CTCTAGTGAA1200
 CAGATGTCAA GAACAAACTA TCAAAGCTTC CACTTTAACAA AACCTGA 1248
 (13) INFORMATION FOR SEQ ID NO:12:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 415 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Ser Gly Met Glu Lys Leu Gln Asn Ala Ser Trp Ile Tyr Gln Gln				
1	5	10	15	
Lys Leu Glu Asp Pro Phe Gln Lys His Leu Asn Ser Thr Glu Glu Tyr				
	20	25	30	
Leu Ala Phe Leu Cys Gly Pro Arg Arg Ser His Phe Phe Leu Pro Val				
	35	40	45	
Ser Val Val Tyr Val Pro Ile Phe Val Val Gly Val Ile Gly Asn Val				
	50	55	60	
Leu Val Cys Leu Val Ile Leu Gln His Gln Ala Met Lys Thr Pro Thr				
30	65	70	75	80
Asn Tyr Tyr Leu Phe Ser Leu Ala Val Ser Asp Leu Leu Val Leu Leu				
	85	90	95	

- 16 -

Leu Gly Met Pro Leu Glu Val Tyr Glu Met Trp Arg Asn Tyr Pro Phe
 100 105 110

Leu Phe Gly Pro Val Gly Cys Tyr Phe Lys Thr Ala Leu Phe Glu Thr
 115 120 125

5 Val Cys Phe Ala Ser Ile Leu Ser Ile Thr Thr Val Ser Val Glu Arg
 130 135 140

Tyr Val Ala Ile Leu His Pro Phe Arg Ala Lys Leu Gln Ser Thr Arg
 145 150 155 160

10 Arg Arg Ala Leu Arg Ile Leu Gly Ile Val Trp Gly Phe Ser Val Leu
 165 170 175

Phe Ser Leu Pro Asn Thr Ser Ile His Gly Ile Lys Phe His Tyr Phe
 180 185 190

Pro Asn Gly Ser Leu Val Pro Gly Ser Ala Thr Cys Thr Val Ile Lys
 195 200 205

15 Pro Met Trp Ile Tyr Asn Phe Ile Ile Gln Val Thr Ser Phe Leu Phe
 210 215 220

Tyr Leu Leu Pro Met Thr Val Ile Ser Val Leu Tyr Tyr Leu Met Ala
 225 230 235 240

20 Leu Arg Leu Lys Lys Asp Lys Ser Leu Glu Ala Asp Glu Gly Asn Ala
 245 250 255

Asn Ile Gln Arg Pro Cys Arg Lys Ser Val Asn Lys Met Leu Phe Val
 260 265 270

Leu Val Leu Val Phe Ala Ile Cys Trp Ala Pro Phe His Ile Asp Arg
 275 280 285

25 Leu Phe Phe Ser Phe Val Glu Glu Trp Ser Glu Ser Leu Ala Ala Val
 290 295 300

Phe Asn Leu Val His Val Val Ser Gly Val Phe Phe Tyr Leu Ser Ser
 305 310 315 320

30 Ala Val Asn Pro Ile Ile Tyr Asn Leu Leu Ser Arg Arg Phe Gln Ala
 325 330 335

Ala Phe Gln Asn Val Ile Ser Ser Phe His Lys Gln Trp His Ser Gln
 340 345 350

His Asp Pro Gln Leu Pro Pro Ala Gln Arg Asn Ile Phe Leu Thr Glu
 355 360 365

35 Cys His Phe Val Glu Leu Thr Glu Asp Ile Gly Pro Gln Phe Pro Cys
 370 375 380

Gln Ser Ser Met His Asn Ser His Leu Pro Thr Ala Leu Ser Ser Glu
 385 390 395 400

- 17 -

Gln	Met	Ser	Arg	Thr	Asn	Tyr	Gln	Ser	Phe	His	Phe	Asn	Lys	Thr
405														410
														415

(14) INFORMATION FOR SEQ ID NO:13:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1173 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

ATGCCAGATA CTAATAGCAC AATCAATTAA TCACTAAGCA CTCGTGTTAC TTTAGCATT 60
 TTTATGTCCT TAGTAGCTTT TGCTATAATG CTAGGAAATG CTTTGGTCAT TTTAGCTTT 120
 GTGGTGGACA AAAACCTTAG ACATCGAAGT AGTTATTTT TTCTTAACCT GGCCATCTCT 180
 GACTTCTTG TGGGTGTGAT CTCCATTCT TTGTACATCC CTCACACGCT GTTCGAATGG 240
 15 GATTTTGGAA AGGAAATCTG TGTATTTGG CTCACTACTG ACTATCTGTT ATGTACAGCA 300
 TCTGTATATA ACATTGTCCT CATCAGCTAT GATCGATACC TGTCACTCTC AAATGCTGTG 360
 TCTTATAGAA CTCAACATAC TGGGGTCTTG AAGATTGTTA CTCTGATGGT GGCGTTTGG 420
 GTGCTGGCCT TCTTAGTGAA TGGGCCAATG ATTCTAGTTT CAGAGTCTTG GAAGGATGAA 480
 GGTAGTGAAT GTGAACCTGG ATTTTTTCG GAATGGTACA TCCTTGCCAT CACATCATTC 540
 20 TTGGAATTCTG TGATCCCAGT CATCTTAGTC GCTTATTTCA ACATGAATAT TTATTGGAGC 600
 CTGTGGAAGC GTGATCATCT CAGTAGGTGC CAAAGCCATC CTGGACTGAC TGCTGTCTCT 660
 TCCAACATCT GTGGACACTC ATTCAAGAGGT AGACTATCTT CAAGGAGATC TCTTCTGCA 720
 TCGACAGAAG TTCCCTGCATC CTTTCATTCA GAGAGACAGA GGAGAAAGAG TAGTCTCATG 780
 TTTCTCTCAA GAACCAAGAT GAATAGCAAT ACAATTGCTT CCAAAATGGG TTCTTCTCC 840
 25 CAATCAGATT CTGTAGCTCT TCACCAAAGG GAACATGTG AACTGCTTAG AGCCAGGAGA 900
 TTAGCCAAGT CACTGGCCAT TCTCTTAGGG GTTTTGCTG TTTGCTGGGC TCCATATTCT 960
 CTGTTCACAA TTGTCCTTTC ATTTTATTCC TCAGCAACAG GTCCTAAATC AGTTGGTAT1020
 AGAATTGCAT TTTGGCTTCA GTGGTTCAAT TCCTTGTCA ATCCTCTTT GTATCCATTG1080
 TGTCACAAGC GCTTCAAAAA GGCTTCTTG AAAATATTTT GTATAAAAAA GCAACCTCTA1140
 30 CCATCACAAAC ACAGTCGGTC AGTATCTTCT TAA

- 18 -

(15) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 390 amino acids
- (B) TYPE: amino acid
- 5 (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

10	Met Pro Asp Thr Asn Ser Thr Ile Asn Leu Ser Leu Ser Thr Arg Val	1	5	10	15
	Thr Leu Ala Phe Phe Met Ser Leu Val Ala Phe Ala Ile Met Leu Gly	20	25	30	
	Asn Ala Leu Val Ile Leu Ala Phe Val Val Asp Lys Asn Leu Arg His	35	40	45	
15	Arg Ser Ser Tyr Phe Phe Leu Asn Leu Ala Ile Ser Asp Phe Phe Val	50	55	60	
	Gly Val Ile Ser Ile Pro Leu Tyr Ile Pro His Thr Leu Phe Glu Trp	65	70	75	80
20	Asp Phe Gly Lys Glu Ile Cys Val Phe Trp Leu Thr Thr Asp Tyr Leu	85	90	95	
	Leu Cys Thr Ala Ser Val Tyr Asn Ile Val Leu Ile Ser Tyr Asp Arg	100	105	110	
	Tyr Leu Ser Val Ser Asn Ala Val Ser Tyr Arg Thr Gln His Thr Gly	115	120	125	
25	Val Leu Lys Ile Val Thr Leu Met Val Ala Val Trp Val Leu Ala Phe	130	135	140	
	Leu Val Asn Gly Pro Met Ile Leu Val Ser Glu Ser Trp Lys Asp Glu	145	150	155	160
30	Gly Ser Glu Cys Glu Pro Gly Phe Phe Ser Glu Trp Tyr Ile Leu Ala	165	170	175	
	Ile Thr Ser Phe Leu Glu Phe Val Ile Pro Val Ile Leu Val Ala Tyr	180	185	190	
	Phe Asn Met Asn Ile Tyr Trp Ser Leu Trp Lys Arg Asp His Leu Ser	195	200	205	
35	Arg Cys Gln Ser His Pro Gly Leu Thr Ala Val Ser Ser Asn Ile Cys	210	215	220	

- 19 -

Gly His Ser Phe Arg Gly Arg Leu Ser Ser Arg Arg Ser Leu Ser Ala
 225 230 235 240
 Ser Thr Glu Val Pro Ala Ser Phe His Ser Glu Arg Gln Arg Arg Lys
 245 250 255
 5 Ser Ser Leu Met Phe Ser Ser Arg Thr Lys Met Asn Ser Asn Thr Ile
 260 265 270
 Ala Ser Lys Met Gly Ser Phe Ser Gln Ser Asp Ser Val Ala Leu His
 275 280 285
 10 Gln Arg Glu His Val Glu Leu Leu Arg Ala Arg Arg Leu Ala Lys Ser
 290 295 300
 Leu Ala Ile Leu Leu Gly Val Phe Ala Val Cys Trp Ala Pro Tyr Ser
 305 310 315 320
 Leu Phe Thr Ile Val Leu Ser Phe Tyr Ser Ser Ala Thr Gly Pro Lys
 325 330 335
 15 Ser Val Trp Tyr Arg Ile Ala Phe Trp Leu Gln Trp Phe Asn Ser Phe
 340 345 350
 Val Asn Pro Leu Leu Tyr Pro Leu Cys His Lys Arg Phe Gln Lys Ala
 355 360 365
 20 Phe Leu Lys Ile Phe Cys Ile Lys Lys Gln Pro Leu Pro Ser Gln His
 370 375 380
 Ser Arg Ser Val Ser Ser
 385 390

(16) INFORMATION FOR SEQ ID NO:15:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1128 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

ATGGCGAACG CGAGCGAGCC GGGTGGCAGC GGCGGCGGCG AGGC GGCCGC CCTGGGCCTC 60
 AAGCTGGCCA CGCTCAGCCT GCTGCTGTGC GTGAGCCTAG CGGGCAACGT GCTGTTCGCG 120
 CTGCTGATCG TGCGGGAGCG CAGCCTGCAC CGCGCCCCGT ACTACCTGCT GCTCGACCTG 180
 TGCCTGGCCG ACGGGCTGCG CGCGCTCGCC TGCCTCCGG CCGTCATGCT GGCGGCGCGG 240
 35 CGTGCAGCGG CCGCGGCGGG GGCGCCGCCG GGCGCGCTGG GCTGCAAGCT GCTCGCCTTC 300

- 20 -

CTGGCCGCGC TCTTCTGCTT CCACGCCGCC TTCTGCTGC TGGCGTGGG CGTCACCCGC 360
 TACCTGGCCA TCGCGCACCA CCGCTTCTAT GCAGAGCGCC TGGCCGGCTG GCCGTGCGCC 420
 GCCATGCTGG TGTGCGCCGC CTGGCGCTG GCGCTGGCCG CGGCCTTCCC GCCAGTGCTG 480
 GACGGCGGTG GCGACGACGA GGACGCGCCG TGCGCCCTGG AGCAGCGGCC CGACGGCGCC 540
 5 CCCGGCGCGC TGGGCTTCCT GCTGCTGCTG GCCGTGGTGG TGGGCGCCAC GCACCTCGTC 600
 TACCTCCGCC TGCTCTTCTT CATCCACGAC CGCCGCAAGA TGCGGCCCGC GCGCCTGGTG 660
 CCCGCCGTCA GCCACGACTG GACCTTCCAC GGCCCGGGCG CCACCGGCCA GGCGGCCGCC 720
 AACTGGACGG CGGGCTTCGG CCGCGGGCCC ACGCCGCCCG CGCTTGTGGG CATCCGGCCC 780
 GCAGGGCCGG GCCGCGGCCGC GCGCCGCCTC CTCGTGCTGG AAGAATTCAA GACGGAGAAG 840
 10 AGGCTGTGCA AGATGTTCTA CGCCGTACAG CTGCTCTTCC TGCTCCTCTG GGGGCCCTAC 900
 GTCGTGGCCA GCTACCTGCG GGTCTGGTG CGGCCCGGGCG CCGTCCCCCA GGCTACCTG 960
 ACGGCCTCCG TGTGGCTGAC CTTCGCGCAG GCCGGCATCA ACCCCGTCGT GTGCTTCCTC 1020
 TTCAACAGGG AGCTGAGGGA CTGCTTCAGG GCCCAGTTCC CCTGCTGCCA GAGCCCCCGG 1080
 ACCACCCAGG CGACCCATCC CTGCGACCTG AAAGGCATTG GTTTATGA 1128

15 (17) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 375 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

20 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Met	Ala	Asn	Ala	Ser	Glu	Pro	Gly	Gly	Ser	Gly	Gly	Gly	Glu	Ala	Ala	
1									10					15		
25	Ala	Leu	Gly	Leu	Lys	Leu	Ala	Thr	Leu	Ser	Leu	Leu	Leu	Cys	Val	Ser
								20					25		30	
	Leu	Ala	Gly	Asn	Val	Leu	Phe	Ala	Leu	Leu	Ile	Val	Arg	Glu	Arg	Ser
								35				40		45		
30	Leu	His	Arg	Ala	Pro	Tyr	Tyr	Leu	Leu	Leu	Asp	Leu	Cys	Leu	Ala	Asp
								50				55		60		
	Gly	Leu	Arg	Ala	Leu	Ala	Cys	Leu	Pro	Ala	Val	Met	Leu	Ala	Ala	Arg
								65				70		75		80

- 21 -

	Arg Ala Ala Ala Ala Ala Gly Ala Pro Pro Gly Ala Leu Gly Cys Lys			
	85	90	95	
	Leu Leu Ala Phe Leu Ala Ala Leu Phe Cys Phe His Ala Ala Phe Leu			
	100	105	110	
5	Leu Leu Gly Val Gly Val Thr Arg Tyr Leu Ala Ile Ala His His Arg			
	115	120	125	
	Phe Tyr Ala Glu Arg Leu Ala Gly Trp Pro Cys Ala Ala Met Leu Val			
	130	135	140	
10	Cys Ala Ala Trp Ala Leu Ala Leu Ala Ala Phe Pro Pro Val Leu			
	145	150	155	160
	Asp Gly Gly Asp Asp Glu Asp Ala Pro Cys Ala Leu Glu Gln Arg			
	165	170	175	
	Pro Asp Gly Ala Pro Gly Ala Leu Gly Phe Leu Leu Leu Leu Ala Val			
	180	185	190	
15	Val Val Gly Ala Thr His Leu Val Tyr Leu Arg Leu Leu Phe Phe Ile			
	195	200	205	
	His Asp Arg Arg Lys Met Arg Pro Ala Arg Leu Val Pro Ala Val Ser			
	210	215	220	
20	His Asp Trp Thr Phe His Gly Pro Gly Ala Thr Gly Gln Ala Ala Ala			
	225	230	235	240
	Asn Trp Thr Ala Gly Phe Gly Arg Gly Pro Thr Pro Pro Ala Leu Val			
	245	250	255	
	Gly Ile Arg Pro Ala Gly Pro Gly Arg Gly Ala Arg Arg Leu Leu Val			
	260	265	270	
25	Leu Glu Glu Phe Lys Thr Glu Lys Arg Leu Cys Lys Met Phe Tyr Ala			
	275	280	285	
	Val Thr Leu Leu Phe Leu Leu Leu Trp Gly Pro Tyr Val Val Ala Ser			
	290	295	300	
30	Tyr Leu Arg Val Leu Val Arg Pro Gly Ala Val Pro Gln Ala Tyr Leu			
	305	310	315	320
	Thr Ala Ser Val Trp Leu Thr Phe Ala Gln Ala Gly Ile Asn Pro Val			
	325	330	335	
	Val Cys Phe Leu Phe Asn Arg Glu Leu Arg Asp Cys Phe Arg Ala Gln			
	340	345	350	
35	Phe Pro Cys Cys Gln Ser Pro Arg Thr Thr Gln Ala Thr His Pro Cys			
	355	360	365	
	Asp Leu Lys Gly Ile Gly Leu			

- 22 -

370

375

(18) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1002 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

10 ATGAACACCA CAGTGATGCA AGGCTTCAAC AGATCTGAGC GGTGCCAG AGACACTCGG 60
 ATAGTACAGC TGGTATTCCC AGCCCTCTAC ACAGTGGTT TCTTGACCGG CATCCTGCTG 120
 AATACTTTGG CTCTGTGGGT GTTGTTAC ATCCCCAGCT CCTCCACCTT CATCATCTAC 180
 CTCAAAAACA CTTGGTGGC CGACTTGATA ATGACACTCA TGCTTCCTT CAAAATCCTC 240
 TCTGACTCAC ACCTGGCACC CTGGCAGCTC AGAGCTTTG TGTGTCGTT TTCTCGGTG 300
 15 ATATTTTATG AGACCATGTA TGTGGGCATC GTGCTGTTAG GGCTCATAGC CTTGACAGA 360
 TTCCTCAAGA TCATCAGACC TTTGAGAAAT ATTTTCTAA AAAAACCTGT TTTGCAAAA 420
 ACGGTCTCAA TCTTCATCTG GTTCTTTTG TTCTTCATCT CCCTGCCAAA TACGATCTG 480
 AGCAACAAAGG AAGAACACCC ATCGTCTGTG AAAAAGTGTG CTTCCCTAAA GGGGCCTCTG 540
 GGGCTGAAAT GGCATCAAAT GGTAAATAAC ATATGCCAGT TTATTTCTG GACTGTTTT 600
 20 ATCCTAATGC TTGTGTTTA TGTGGTTATT GCAAAAAAAG TATATGATT C TTATAGAAAG 660
 TCCAAAAGTA AGGACAGAAA AAACAACAAA AAGCTGGAAG GCAAAGTATT TGTTGTCGTG 720
 GCTGTCTTCT TTGTGTTTT TGCTCCATT CATTGCCA GAGTTCCATA TACTCACAGT 780
 CAAACCAACA ATAAGACTGA CTGTAGACTG CAAAATCAAC TGTTTATTGC TAAAGAAACA 840
 ACTCTCTTT TGGCAGCAAC TAACATTTGT ATGGATCCCT TAATATACAT ATTCTTATGT 900
 25 AAAAATTCA CAGAAAAGCT ACCATGTATG CAAGGGAGAA AGACCCACAGC ATCAAGCCAA 960
 GAAAATCATA GCAGTCAGAC AGACAAACATA ACCTTAGGCT GA 1002

(19) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 333 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:

30

- 23 -

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

5	Met Asn Thr Thr Val Met Gln Gly Phe Asn Arg Ser Glu Arg Cys Pro	1	5	10	15
	Arg Asp Thr Arg Ile Val Gln Leu Val Phe Pro Ala Leu Tyr Thr Val	20	25	30	
	Val Phe Leu Thr Gly Ile Leu Leu Asn Thr Leu Ala Leu Trp Val Phe	35	40	45	
10	Val His Ile Pro Ser Ser Ser Thr Phe Ile Ile Tyr Leu Lys Asn Thr	50	55	60	
	Leu Val Ala Asp Leu Ile Met Thr Leu Met Leu Pro Phe Lys Ile Leu	65	70	75	80
15	Ser Asp Ser His Leu Ala Pro Trp Gln Leu Arg Ala Phe Val Cys Arg	85	90	95	
	Phe Ser Ser Val Ile Phe Tyr Glu Thr Met Tyr Val Gly Ile Val Leu	100	105	110	
	Leu Gly Leu Ile Ala Phe Asp Arg Phe Leu Lys Ile Ile Arg Pro Leu	115	120	125	
20	Arg Asn Ile Phe Leu Lys Lys Pro Val Phe Ala Lys Thr Val Ser Ile	130	135	140	
	Phe Ile Trp Phe Phe Leu Phe Phe Ile Ser Leu Pro Asn Thr Ile Leu	145	150	155	160
25	Ser Asn Lys Glu Ala Thr Pro Ser Ser Val Lys Lys Cys Ala Ser Leu	165	170	175	
	Lys Gly Pro Leu Gly Leu Lys Trp His Gln Met Val Asn Asn Ile Cys	180	185	190	
	Gln Phe Ile Phe Trp Thr Val Phe Ile Leu Met Leu Val Phe Tyr Val	195	200	205	
30	Val Ile Ala Lys Lys Val Tyr Asp Ser Tyr Arg Lys Ser Lys Ser Lys	210	215	220	
	Asp Arg Lys Asn Asn Lys Lys Leu Glu Gly Lys Val Phe Val Val Val	225	230	235	240
35	Ala Val Phe Phe Val Cys Phe Ala Pro Phe His Phe Ala Arg Val Pro	245	250	255	

- 24 -

Tyr Thr His Ser Gln Thr Asn Asn Lys Thr Asp Cys Arg Leu Gln Asn
 260 265 270

Gln Leu Phe Ile Ala Lys Glu Thr Thr Leu Phe Leu Ala Ala Thr Asn
 275 280 285

5 Ile Cys Met Asp Pro Leu Ile Tyr Ile Phe Leu Cys Lys Lys Phe Thr
 290 295 300

Glu Lys Leu Pro Cys Met Gln Gly Arg Lys Thr Thr Ala Ser Ser Gln
 305 310 315 320

10 Glu Asn His Ser Ser Gln Thr Asp Asn Ile Thr Leu Gly
 325 330

(20) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1122 base pairs
 (B) TYPE: nucleic acid
 15 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

ATGGCCAACA CTACCGGAGA GCCTGAGGAG GTGAGCGGCG CTCTGTCCCC ACCGTCCGCA 60

20 TCAGCTTATG TGAAGCTGGT ACTGCTGGG A CTGATTATGT GCGTGAGCCT GGCGGGTAAC 120

GCCATCTTGT CCCTGCTGGT GCTCAAGGAG CGTGCCCTGC ACAAGGCTCC TTACTACTTC 180

CTGCTGGACC TGTGCCTGGC CGATGGCATA CGCTCTGCCG TCTGCTTCCC CTTTGTGCTG 240

GCTTCTGTGC GCCACGGCTC TTCATGGACC TTCAGTGCAC TCAGCTGCAA GATTGTGGCC 300

TTTATGGCCG TGCTCTTTG CTTCCATGCG GCCTTCATGC TGTTCTGCAT CAGCGTCACC 360

25 CGCTACATGG CCATCGCCCA CCACCGCTTC TACGCCAAGC GCATGACACT CTGGACATGC 420

GCGGCTGTCA TCTGCATGGC CTGGACCCCTG TCTGTGGCCA TGGCCTTCCC ACCTGTCTT 480

GACGTGGGCA CCTACAAGTT TATTGGGAG GAGGACCAAGT GCATCTTGA GCATCGCTAC 540

TTCAAGGCCA ATGACACGCT GGGCTTCATG CTTATGTTGG CTGTGCTCAT GGCAGCTACC 600

CATGCTGTCT ACGGCAAGCT GCTCCTCTTC GAGTATCGTC ACCGCAAGAT GAAGCCAGTG 660

30 CAGATGGTGC CAGCCATCAG CCAGAACTGG ACATTCCATG GTCCCGGGGC CACCGGCCAG 720

GCTGCTGCCA ACTGGATCGC CGGCTTGCG CGTGGGCCA TGCCACCAAC CCTGCTGGGT 780

ATCCGGCAGA ATGGGCATGC AGCCAGCCGG CGGCTACTGG GCATGGACGA GGTCAAGGGT 840

- 25 -

GAAAAGCAGC TGGGCCGCAT GTTCTACGCG ATCACACTGC TCTTTCTGCT CCTCTGGTCA 900
 CCCTACATCG TGGCCTGCTA CTGGCGAGTG TTTGTGAAAG CCTGTGCTGT GCCCCACCGC 960
 TACCTGGCCA CTGCTGTTG GATGAGCTTC GCCCAGGCTG CCGTCAACCC AATTGTCTGC1020
 TTCCTGCTCA ACAAGGACCT CAAGAAAGTGC CTGACCACTC ACGCCCCCTG CTGGGGCAC1080
 5 GGAGGTGCCCG CGGCTCCCAG AGAACCTAC TGTGTATGT GA 1122

(21) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 373 amino acids
- (B) TYPE: amino acid
- 10 (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

15 Met Ala Asn Thr Thr Gly Glu Pro Glu Glu Val Ser Gly Ala Leu Ser
 1 5 10 15
 Pro Pro Ser Ala Ser Ala Tyr Val Lys Leu Val Leu Leu Gly Leu Ile
 20 25 30
 Met Cys Val Ser Leu Ala Gly Asn Ala Ile Leu Ser Leu Leu Val Leu
 35 40 45
 20 Lys Glu Arg Ala Leu His Lys Ala Pro Tyr Tyr Phe Leu Leu Asp Leu
 50 55 60
 Cys Leu Ala Asp Gly Ile Arg Ser Ala Val Cys Phe Pro Phe Val Leu
 65 70 75 80
 25 Ala Ser Val Arg His Gly Ser Ser Trp Thr Phe Ser Ala Leu Ser Cys
 85 90 95
 Lys Ile Val Ala Phe Met Ala Val Leu Phe Cys Phe His Ala Ala Phe
 100 105 110
 Met Leu Phe Cys Ile Ser Val Thr Arg Tyr Met Ala Ile Ala His His
 115 120 125
 30 Arg Phe Tyr Ala Lys Arg Met Thr Leu Trp Thr Cys Ala Ala Val Ile
 130 135 140
 Cys Met Ala Trp Thr Leu Ser Val Ala Met Ala Phe Pro Pro Val Phe
 145 150 155 160
 35 Asp Val Gly Thr Tyr Lys Phe Ile Arg Glu Glu Asp Gln Cys Ile Phe
 165 170 175

- 26 -

Glu His Arg Tyr Phe Lys Ala Asn Asp Thr Leu Gly Phe Met Leu Met
 180 185 190
 Leu Ala Val Leu Met Ala Ala Thr His Ala Val Tyr Gly Lys Leu Leu
 195 200 205
 5 Leu Phe Glu Tyr Arg His Arg Lys Met Lys Pro Val Gln Met Val Pro
 210 215 220
 Ala Ile Ser Gln Asn Trp Thr Phe His Gly Pro Gly Ala Thr Gly Gln
 225 230 235 240
 Ala Ala Ala Asn Trp Ile Ala Gly Phe Gly Arg Gly Pro Met Pro Pro
 10 245 250 255
 Thr Leu Leu Gly Ile Arg Gln Asn Gly His Ala Ala Ser Arg Arg Leu
 260 265 270
 Leu Gly Met Asp Glu Val Lys Gly Glu Lys Gln Leu Gly Arg Met Phe
 275 280 285
 15 Tyr Ala Ile Thr Leu Leu Phe Leu Leu Leu Trp Ser Pro Tyr Ile Val
 290 295 300
 Ala Cys Tyr Trp Arg Val Phe Val Lys Ala Cys Ala Val Pro His Arg
 305 310 315 320
 Tyr Leu Ala Thr Ala Val Trp Met Ser Phe Ala Gln Ala Ala Val Asn
 20 325 330 335
 Pro Ile Val Cys Phe Leu Leu Asn Lys Asp Leu Lys Lys Cys Leu Thr
 340 345 350
 Thr His Ala Pro Cys Trp Gly Thr Gly Gly Ala Pro Ala Pro Arg Glu
 355 360 365
 25 Pro Tyr Cys Val Met
 370

(22) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

30 (A) LENGTH: 1053 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

35 ATGGCTTG AACAGAACCA GTCAACAGAT TATTATTATG AGGAAAATGA AATGAATGGC 60
 ACTTATGACT ACAGTCAATA TGAATTGATC TGTATCAAAG AAGATGTCAG AGAATTGCA 120

- 27 -

AAAGTTTCC TCCCTGTATT CCTCACAAATA GCTTCGTCA TTGGACTTGC AGGCAATTCC 180
 ATGGTAGTGG CAATTTATGC CTATTACAAG AAACAGAGAA CCAAAACAGA TGTGTACATC 240
 CTGAATTTGG CTGTAGCAGA TTTACTCCTT CTATTCACTC TGCCTTTTG GGCTGTTAAT 300
 GCAGTTCATG GGTGGGTTTT AGGGAAAATA ATGTGCAAAA TAACTTCAGC CTTGTACACA 360
 5 CTAAACTTG TCTCTGGAAT GCAGTTCTG GCTTGCATCA GCATAGACAG ATATGTGGCA 420
 GTAACTAATG TCCCCAGCCA ATCAGGAGTG GGAAAACCAT GCTGGATCAT CTGTTCTGT 480
 GTCTGGATGG CTGCCATCTT GCTGAGCATA CCCCAGCTGG TTTTTATAC AGTAAATGAC 540
 AATGCTAGGT GCATTCCCAT TTTCCCCGC TACCTAGGAA CATCAATGAA AGCATTGATT 600
 CAAATGCTAG AGATCTGCAT TGGATTTGTA GTACCCTTTC TTATTATGGG GGTGTGCTAC 660
 10 TTTATCACGG CAAGGACACT CATGAAGATG CCAAACATTA AAATATCTCG ACCCCTAAAAA 720
 GTTCTGCTCA CAGTCGTTAT AGTTTCATT GTCACTCAAC TGCCTTATAA CATTGTCAAG 780
 TTCTGCCGAG CCATAGACAT CATCTACTCC CTGATCACCA GCTGCAACAT GAGCAAACGC 840
 ATGGACATCG CCATCCAAGT CACAGAAAGC ATTGCACTCT TTCACAGCTG CCTCAACCCA 900
 ATCCTTATG TTTTATGGG AGCATCTTC AAAAACTACG TTATGAAAGT GGCCAAGAAA 960
 15 TATGGGTCT GGAGAAGACA GAGACAAAGT GTGGAGGAGT TTCCCTTTGA TTCTGAGGGT1020
 CCTACAGAGC CAACCAAGTAC TTTAGCATT TAA

1053

(23) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 350 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

25 Met Ala Leu Glu Gln Asn Gln Ser Thr Asp Tyr Tyr Tyr Glu Glu Asn
 1 5 10 15

Glu Met Asn Gly Thr Tyr Asp Tyr Ser Gln Tyr Glu Leu Ile Cys Ile
 20 25 30

30 Lys Glu Asp Val Arg Glu Phe Ala Lys Val Phe Leu Pro Val Phe Leu
 35 40 45

Thr Ile Ala Phe Val Ile Gly Leu Ala Gly Asn Ser Met Val Val Ala

- 28 -

	50	55	60
	Ile Tyr Ala Tyr Tyr Lys Lys Gln Arg Thr Lys Thr Asp Val Tyr Ile		
65	65	70	75
			80
5	Leu Asn Leu Ala Val Ala Asp Leu Leu Leu Leu Phe Thr Leu Pro Phe		
	85	90	95
	Trp Ala Val Asn Ala Val His Gly Trp Val Leu Gly Lys Ile Met Cys		
	100	105	110
	Lys Ile Thr Ser Ala Leu Tyr Thr Leu Asn Phe Val Ser Gly Met Gln		
	115	120	125
10	Phe Leu Ala Cys Ile Ser Ile Asp Arg Tyr Val Ala Val Thr Asn Val		
	130	135	140
	Pro Ser Gln Ser Gly Val Gly Lys Pro Cys Trp Ile Ile Cys Phe Cys		
	145	150	155
			160
15	Val Trp Met Ala Ala Ile Leu Leu Ser Ile Pro Gln Leu Val Phe Tyr		
	165	170	175
	Thr Val Asn Asp Asn Ala Arg Cys Ile Pro Ile Phe Pro Arg Tyr Leu		
	180	185	190
	Gly Thr Ser Met Lys Ala Leu Ile Gln Met Leu Glu Ile Cys Ile Gly		
	195	200	205
20	Phe Val Val Pro Phe Leu Ile Met Gly Val Cys Tyr Phe Ile Thr Ala		
	210	215	220
	Arg Thr Leu Met Lys Met Pro Asn Ile Lys Ile Ser Arg Pro Leu Lys		
	225	230	235
			240
25	Val Leu Leu Thr Val Val Ile Val Phe Ile Val Thr Gln Leu Pro Tyr		
	245	250	255
	Asn Ile Val Lys Phe Cys Arg Ala Ile Asp Ile Ile Tyr Ser Leu Ile		
	260	265	270
	Thr Ser Cys Asn Met Ser Lys Arg Met Asp Ile Ala Ile Gln Val Thr		
	275	280	285
30	Glu Ser Ile Ala Leu Phe His Ser Cys Leu Asn Pro Ile Leu Tyr Val		
	290	295	300
	Phe Met Gly Ala Ser Phe Lys Asn Tyr Val Met Lys Val Ala Lys Lys		
	305	310	315
			320
35	Tyr Gly Ser Trp Arg Arg Gln Arg Gln Ser Val Glu Glu Phe Pro Phe		
	325	330	335
	Asp Ser Glu Gly Pro Thr Glu Pro Thr Ser Thr Phe Ser Ile		
	340	345	350

- 29 -

(24) INFORMATION FOR SEQ ID NO:23:

5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1116 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```

ATGCCAGGAA ACGCCACCCC AGTGACCACC ACTGCCCGT GGGCCTCCCT GGGCCTCTCC 60
10 GCCAAGACCT GCAACAAACGT GTCCTTCGAA GAGAGCAGGA TAGTCCTGGT CGTGGTGTAC 120
  AGCGCGGTGT GCACGCTGGG GGTGCCGCC AACTGCCTGA CTGCGTGGCT GGCGCTGCTG 180
  CAGGTACTGC AGGGCAACGT GCTGGCCGTC TACCTGCTCT GCCTGGCACT CTGCGAACTG 240
  CTGTACACAG GCACGCTGCC ACTCTGGTC ATCTATATCC GCAACCAGCA CCGCTGGACC 300
  CTAGGCCTGC TGGCCTCGAA GGTGACCGCC TACATCTTCT TCTGCAACAT CTACGTCAGC 360
15 ATCCTCTTCC TGTGCTGCAT CTCCTGCGAC CGCTTCGTGG CCGTGGTGT A CGCGCTGGAG 420
  AGTCGGGGCC GCCGCCGCC GAGGACCGCC ATCCTCATCT CCGCCTGCAT CTTCATCCTC 480
  GTCGGGATCG TTCACTACCC GGTGTTCCAG ACGGAAGACA AGGAGACCTG CTTTGACATG 540
  CTGCAGATGG ACAGCAGGAT TGCCGGGTAC TACTACGCCA GGTTCACCGT TGGCTTGCC 600
  ATCCCTCTCT CCATCATCGC CTTCACCAAC CACCGGATT TCAGGAGCAT CAAGCAGAGC 660
20 ATGGGCTTAA GCGCTGCCA GAAGGCCAAG GTGAAGCACT CGGCCATCGC GGTGGTTGTC 720
  ATCTTCCTAG TCTGCTTCGC CCCGTACCAAC CTGGTTCTCC TCGTCAAAGC CGCTGCCTTT 780
  TCCTACTACA GAGGAGACAG GAACGCCATG TGCAGCTTGG AGGAAAGGCT GTACACAGCC 840
  TCTGTGGTGT TTCTGTGCCT GTCCACGGTG AACGGCGTGG CTGACCCCAT TATCTACGTG 900
  CTGGCCACGG ACCATTCCCG CCAAGAAGTG TCCAGAATCC ATAAGGGGTG GAAAGAGTGG 960
25 TCCATGAAGA CAGACGTAC CAGGCTCACC CACAGCAGGG ACACCGAGGA GCTGCAGTCG 1020
  CCCGTGGCCC TTGCAGACCA CTACACCTTC TCCAGGCCCG TGCACCCACC AGGGTCACCA 1080
  TGCCCTGCAA AGAGGCTGAT TGAGGGAGTCC TGCTGA
  1116

```

(25) INFORMATION FOR SEQ ID NO:24:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 371 amino acids

- 30 -

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Met Pro Gly Asn Ala Thr Pro Val Thr Thr Thr Ala Pro Trp Ala Ser
 1 5 10 15

Leu Gly Leu Ser Ala Lys Thr Cys Asn Asn Val Ser Phe Glu Glu Ser
 20 25 30

10 Arg Ile Val Leu Val Val Val Tyr Ser Ala Val Cys Thr Leu Gly Val
 35 40 45

Pro Ala Asn Cys Leu Thr Ala Trp Leu Ala Leu Leu Gln Val Leu Gln
 50 55 60

Gly Asn Val Leu Ala Val Tyr Leu Leu Cys Leu Ala Leu Cys Glu Leu
 65 70 75 80

Leu Tyr Thr Gly Thr Leu Pro Leu Trp Val Ile Tyr Ile Arg Asn Gln
 85 90 95

His Arg Trp Thr Leu Gly Leu Leu Ala Ser Lys Val Thr Ala Tyr Ile
 100 105 110

20 Phe Phe Cys Asn Ile Tyr Val Ser Ile Leu Phe Leu Cys Cys Ile Ser
 115 120 125

Cys Asp Arg Phe Val Ala Val Val Tyr Ala Leu Glu Ser Arg Gly Arg
 130 135 140

25 Arg Arg Arg Arg Thr Ala Ile Leu Ile Ser Ala Cys Ile Phe Ile Leu
 145 150 155 160

Val Gly Ile Val His Tyr Pro Val Phe Gln Thr Glu Asp Lys Glu Thr
 165 170 175

Cys Phe Asp Met Leu Gln Met Asp Ser Arg Ile Ala Gly Tyr Tyr Tyr
 180 185 190

30 Ala Arg Phe Thr Val Gly Phe Ala Ile Pro Leu Ser Ile Ile Ala Phe
 195 200 205

Thr Asn His Arg Ile Phe Arg Ser Ile Lys Gln Ser Met Gly Leu Ser
 210 215 220

35 Ala Ala Gln Lys Ala Lys Val Lys His Ser Ala Ile Ala Val Val Val
 225 230 235 240

- 31 -

	Ile Phe Leu Val Cys Phe Ala Pro Tyr His Leu Val Leu Leu Val Lys			
	245	250	255	
	Ala Ala Ala Phe Ser Tyr Tyr Arg Gly Asp Arg Asn Ala Met Cys Gly			
	260	265	270	
5	Leu Glu Glu Arg Leu Tyr Thr Ala Ser Val Val Phe Leu Cys Leu Ser			
	275	280	285	
	Thr Val Asn Gly Val Ala Asp Pro Ile Ile Tyr Val Leu Ala Thr Asp			
	290	295	300	
10	His Ser Arg Gln Glu Val Ser Arg Ile His Lys Gly Trp Lys Glu Trp			
	305	310	315	320
	Ser Met Lys Thr Asp Val Thr Arg Leu Thr His Ser Arg Asp Thr Glu			
	325	330	335	
	Glu Leu Gln Ser Pro Val Ala Leu Ala Asp His Tyr Thr Phe Ser Arg			
	340	345	350	
15	Pro Val His Pro Pro Gly Ser Pro Cys Pro Ala Lys Arg Leu Ile Glu			
	355	360	365	
	Glu Ser Cys			
	370			

(26) INFORMATION FOR SEQ ID NO:25:

20 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1113 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

```

ATGGCGAACT ATAGCCATGC AGCTGACAAC ATTTGCAAA ATCTCTGCC TCTAACAGCC 60
TTTCTGAAAC TGACTTCCTT GGGTTTCATA ATAGGAGTCA GCGTGGTGGG CAACCTCCTG 120
ATCTCCATTG TGCTAGTGAA AGATAAGACC TTGCATAGAG CACCTTACTA CTTCTGTTG 180
30 GATCTTGCT GTTCAGATAT CCTCAGATCT GCAATTGTT TCCCATTGT GTTCAACTCT 240
GTCAAAAATG GCTCTACCTG GACTTATGGG ACTCTGACTT GCAAAGTGAT TGCCTTCTG 300
GGGGTTTGT CCTGTTCCA CACTGCTTC ATGCTCTTCT GCATCAGTGT CACCAAGATAC 360
TTAGCTATCG CCCATCACCG CTTCTATACA AAGAGGGCTGA CCTTTGGAC GTGTCTGGCT 420
GTGATCTGTA TGGTGTGGAC TCTGTCTGTG GCCATGGCAT TTCCCCGGT TTTAGACGTG 480

```

- 32 -

GGCACTTACT CATTCAATTAG GGAGGAAGAT CAATGCACCT TCCAACACCG CTCCTTCAGG 540
 GCTAATGATT CCTTAGGATT TATGCTGCTT CTTGCTCTCA TCCTCCTAGC CACACAGCTT 600
 GTCTACCTCA AGCTGATATT TTTCGTCCAC GATCGAAGAA AAATGAAGCC AGTCCAGTTT 660
 GTAGCAGCAG TCAGCCAGAA CTGGACTTT CATGGTCCTG GAGCCAGTGG CCAGGCAGCT 720
 5 GCCAATTGGC TAGCAGGATT TGGAAGGGT CCCACACCAAC CCACCTTGCT GGGCATCAGG 780
 CAAAATGCAA ACACCACAGG CAGAAGAAGG CTATTGGTCT TAGACGAGTT CAAAATGGAG 840
 AAAAGAATCA GCAGAATGTT CTATATAATG ACTTTTCTGT TTCTAACCTT GTGGGGCCCC 900
 TACCTGGTGG CCTGTTATTG GAGAGTTTT GCAAGAGGGC CTGTAGTACC AGGGGGATTT 960
 CTAACAGCTG CTGTCTGGAT GAGTTTGCC CAAGCAGGAA TCAATCCTT TGTCTGCATT1020
 10 TTCTCAAACA GGGAGCTGAG GCGCTGTTTC AGCACAAACCC TTCTTACTG CAGAAAATCC1080
 AGGTTACCAA GGGAACCTTA CTGTGTTATA TGA 1113

(27) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 370 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

20 Met Ala Asn Tyr Ser His Ala Ala Asp Asn Ile Leu Gln Asn Leu Ser
 1 5 10 15
 Pro Leu Thr Ala Phe Leu Lys Leu Thr Ser Leu Gly Phe Ile Ile Gly
 20 25 30
 Val Ser Val Val Gly Asn Leu Leu Ile Ser Ile Leu Leu Val Lys Asp
 25 35 40 45
 Lys Thr Leu His Arg Ala Pro Tyr Tyr Phe Leu Leu Asp Leu Cys Cys
 50 55 60
 Ser Asp Ile Leu Arg Ser Ala Ile Cys Phe Pro Phe Val Phe Asn Ser
 65 70 75 80
 30 Val Lys Asn Gly Ser Thr Trp Thr Tyr Gly Thr Leu Thr Cys Lys Val
 85 90 95
 Ile Ala Phe Leu Gly Val Leu Ser Cys Phe His Thr Ala Phe Met Leu

- 33 -

	100	105	110
	Phe Cys Ile Ser Val Thr Arg Tyr Leu Ala Ile Ala His His Arg Phe		
	115	120	125
5	Tyr Thr Lys Arg Leu Thr Phe Trp Thr Cys Leu Ala Val Ile Cys Met		
	130	135	140
	Val Trp Thr Leu Ser Val Ala Met Ala Phe Pro Pro Val Leu Asp Val		
	145	150	155
	Gly Thr Tyr Ser Phe Ile Arg Glu Glu Asp Gln Cys Thr Phe Gln His		
	165	170	175
10	Arg Ser Phe Arg Ala Asn Asp Ser Leu Gly Phe Met Leu Leu Leu Ala		
	180	185	190
	Leu Ile Leu Leu Ala Thr Gln Leu Val Tyr Leu Lys Leu Ile Phe Phe		
	195	200	205
15	Val His Asp Arg Arg Lys Met Lys Pro Val Gln Phe Val Ala Ala Val		
	210	215	220
	Ser Gln Asn Trp Thr Phe His Gly Pro Gly Ala Ser Gly Gln Ala Ala		
	225	230	235
	Ala Asn Trp Leu Ala Gly Phe Gly Arg Gly Pro Thr Pro Pro Thr Leu		
	245	250	255
20	Leu Gly Ile Arg Gln Asn Ala Asn Thr Thr Gly Arg Arg Arg Leu Leu		
	260	265	270
	Val Leu Asp Glu Phe Lys Met Glu Lys Arg Ile Ser Arg Met Phe Tyr		
	275	280	285
25	Ile Met Thr Phe Leu Phe Leu Thr Leu Trp Gly Pro Tyr Leu Val Ala		
	290	295	300
	Cys Tyr Trp Arg Val Phe Ala Arg Gly Pro Val Val Pro Gly Gly Phe		
	305	310	315
	Leu Thr Ala Ala Val Trp Met Ser Phe Ala Gln Ala Gly Ile Asn Pro		
	325	330	335
30	Phe Val Cys Ile Phe Ser Asn Arg Glu Leu Arg Arg Cys Phe Ser Thr		
	340	345	350
	Thr Leu Leu Tyr Cys Arg Lys Ser Arg Leu Pro Arg Glu Pro Tyr Cys		
	355	360	365
35	Val Ile		
	370		

(28) INFORMATION FOR SEQ ID NO:27:

- 34 -

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1080 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- 5 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

ATGCAGGTCC CGAACAGCAC CGGCCGGAC AACGCGACGC TGCAGATGCT GCGGAACCCG 60
 GCGATCGCGG TGGCCCTGCC CGTGGTGTAC TCGCTGGTGG CGGCGGTCAG CATCCCGGC 120
 10 AACCTCTTCT CTCTGTGGGT GCTGTGCCGG CGCATGGGC CCAGATCCCC GTCGGTCATC 180
 TTCATGATCA ACCTGAGCGT CACGGACCTG ATGCTGGCCA GCGTGTGTC 240
 TACTACCATT GCAACCGCCA CCACTGGTA TTCGGGGTGC TGCTTGCAA CGTGGTGACC 300
 GTGGCCTTT ACGCAAACAT GTATTCCAGC ATCCTCACCA TGACCTGTAT CAGCGTGGAG 360
 CGCTTCCTGG GGGCCTGTA CCCGCTCAGC TCCAAGCGCT GGCGCCGCCG TCGTTACGCG 420
 15 GTGGCCGCGT GTGCAGGGAC CTGGCTGCTG CTCCTGACCG CCCTGTGCCG GCTGGCGCGC 480
 ACCGATCTCA CCTACCCGGT GCACGCCCTG GGCATCATCA CCTGCTTCGA CGTCCTCAAG 540
 TGGACGATGC TCCCCAGCGT GGCCATGTGG GCCGTGTTCC TCTTCACCAT CTTCATCCTG 600
 CTGTTCTCA TCCCCTCGT GATCACCGTG GCTTGTACA CGGCCACCAT CCTCAAGCTG 660
 TTGCGCACGG AGGAGGCGCA CGGCCGGGAG CAGCGGAGGC GCGCGGTGGG CCTGGCCGCG 720
 20 GTGGTCTTGC TGGCCTTGT CACCTGCTTC GCCCCCAACA ACTTCGTGCT CCTGGCGCAC 780
 ATCGTGAGCC GCCTGTTCTA CGGCAAGAGC TACTACCACG TGTACAAGCT CACGCTGTGT 840
 CTCAGCTGCC TCAACAACTG TCTGGACCCG TTTGTTATT ACTTTGCGTC CCGGGAATTG 900
 CAGCTGCGCC TGGGGAATA TTTGGGCTGC CGCCGGGTGC CCAGAGACAC CCTGGACACG 960
 CGCCGCGAGA GCCTCTTCTC CGCCAGGACC ACGTCCGTGC GCTCCGAGGC CGGTGCGCAC 1020
 25 CCTGAAGGGA TGGAGGGAGC CACCAGGCC GGCTCCAGA GGCAGGAGAG TGTGTTCTGA 1080

(29) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 359 amino acids
- (B) TYPE: amino acid
- 30 (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant

- 35 -

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

	Met Gln Val Pro Asn Ser Thr Gly Pro Asp Asn Ala Thr Leu Gln Met			
1	5	10	15	
5	Leu Arg Asn Pro Ala Ile Ala Val Ala Leu Pro Val Val Tyr Ser Leu			
	20	25	30	
	Val Ala Ala Val Ser Ile Pro Gly Asn Leu Phe Ser Leu Trp Val Leu			
	35	40	45	
10	Cys Arg Arg Met Gly Pro Arg Ser Pro Ser Val Ile Phe Met Ile Asn			
	50	55	60	
	Leu Ser Val Thr Asp Leu Met Leu Ala Ser Val Leu Pro Phe Gln Ile			
	65	70	75	80
	Tyr Tyr His Cys Asn Arg His His Trp Val Phe Gly Val Leu Leu Cys			
	85	90	95	
15	Asn Val Val Thr Val Ala Phe Tyr Ala Asn Met Tyr Ser Ser Ile Leu			
	100	105	110	
	Thr Met Thr Cys Ile Ser Val Glu Arg Phe Leu Gly Val Leu Tyr Pro			
	115	120	125	
20	Leu Ser Ser Lys Arg Trp Arg Arg Arg Tyr Ala Val Ala Ala Cys			
	130	135	140	
	Ala Gly Thr Trp Leu Leu Leu Thr Ala Leu Cys Pro Leu Ala Arg			
	145	150	155	160
	Thr Asp Leu Thr Tyr Pro Val His Ala Leu Gly Ile Ile Thr Cys Phe			
	165	170	175	
25	Asp Val Leu Lys Trp Thr Met Leu Pro Ser Val Ala Met Trp Ala Val			
	180	185	190	
	Phe Leu Phe Thr Ile Phe Ile Leu Leu Phe Leu Ile Pro Phe Val Ile			
	195	200	205	
30	Thr Val Ala Cys Tyr Thr Ala Thr Ile Leu Lys Leu Leu Arg Thr Glu			
	210	215	220	
	Glu Ala His Gly Arg Glu Gln Arg Arg Arg Ala Val Gly Leu Ala Ala			
	225	230	235	240
	Val Val Leu Leu Ala Phe Val Thr Cys Phe Ala Pro Asn Asn Phe Val			
	245	250	255	
35	Leu Leu Ala His Ile Val Ser Arg Leu Phe Tyr Gly Lys Ser Tyr Tyr			
	260	265	270	

- 36 -

	His	Val	Tyr	Lys	Leu	Thr	Leu	Cys	Leu	Ser	Cys	Leu	Asn	Asn	Cys	Leu
				275				280				285				
	Asp	Pro	Phe	Val	Tyr	Tyr	Phe	Ala	Ser	Arg	Glu	Phe	Gln	Leu	Arg	Leu
				290			295				300					
5	Arg	Glu	Tyr	Leu	Gly	Cys	Arg	Arg	Val	Pro	Arg	Asp	Thr	Leu	Asp	Thr
				305			310			315		320				
	Arg	Arg	Glu	Ser	Leu	Phe	Ser	Ala	Arg	Thr	Thr	Ser	Val	Arg	Ser	Glu
				325			330			335						
10	Ala	Gly	Ala	His	Pro	Glu	Gly	Met	Glu	Gly	Ala	Thr	Arg	Pro	Gly	Leu
				340			345			350						
	Gln	Arg	Gln	Glu	Ser	Val	Phe									
				355												

(30) INFORMATION FOR SEQ ID NO:29:

	(i) SEQUENCE CHARACTERISTICS:
15	(A) LENGTH: 1503 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: single
	(D) TOPOLOGY: linear
	(ii) MOLECULE TYPE: DNA (genomic)
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:
	ATGGAGCGTC CCTGGGAGGA CAGCCCAGGC CCGGAGGGGG CAGCTGAGGG CTCGCCTGTG 60
	CCAGTCGCCG CGGGGGCGCG CTCCGGTGCC GCGGCGAGTG GCACAGGCTG GCAGCCATGG 120
	GCTGAGTGCC CGGGACCAA GGGGAGGGGG CAACTGCTGG CGACCGCCGG CCCTTGCGT 180
	CGCTGGCCCG CCCCCTCGCC TGCCAGCTCC AGCCCCGCC CCGGAGCGGC GTCCGCTCAC 240
25	TCGGTTCAAG GCAGCGCGAC TGCGGGTGGC GCACGACCAG GGCGCAGACC TTGGGGCGCG 300
	CGGCCCATGG AGTCGGGGCT GCTGCGGCCG GCGCCGGTGA GCGAGGTCAT CGTCCTGCAT 360
	TACAACTACA CCGGCAAGCT CCGCGGTGCG AGCTACCAGC CGGGTGCCGG CCTGCGCGCC 420
	GACGCCGTGG TGTGCCTGGC GGTGTGCGCC TTCATCGTGC TAGAGAATCT AGCCGTGTTG 480
	TTGGTGCTCG GACGCCACCC GCGCTTCCAC GCTCCCATGT TCCTGCTCCT GGGCAGCCTC 540
30	ACGTTGTCGG ATCTGCTGGC AGGCGCCGCC TACGCCGCCA ACATCCTACT GTCGGGGCCG 600
	CTCACGCTGA AACTGTCCCC CGCGCTCTGG TTCGCACGGG AGGGAGGCCGT CTTCGTGGCA 660
	CTCACTGCGT CCGTGCTGAG CCTCCTGGCC ATCGCGCTGG AGCGCAGCCT CACCATGGCG 720

- 37 -

CGCAGGGGGC CCGCGCCCGT CTCCAGTCGG GGGCGCACGC TGGCGATGGC AGCCGCGGCC 780
 TGGGGCGTGT CGCTGCTCCT CGGGCTCCTG CCAGCGCTGG GCTGGAATTG CCTGGGTCGC 840
 CTGGACGCTT GCTCCACTGT CTTGCCGCTC TACGCCAAGG CCTACGTGCT CTTCTGCGTG 900
 CTCGCCTTCG TGGGCATCCT GGCGCGATC TGTGCACTCT ACGCGCGCAT CTACTGCCAG 960
 5 GTACGCGCCA ACGCGCGCG CCTGCCGGCA CGGCCCGGGA CTGCGGGGAC CACCTCGACC1020
 CGGGCGCGTC GCAAGCCGCG CTCTCTGGCC TTGCTGCGCA CGCTCAGCGT GGTGCTCCTG1080
 GCCTTTGTGG CATGTTGGGG CCCCCTCTTC CTGCTGCTGT TGCTCGACGT GGC GTGCCCG1140
 GCGCGCACCT GTCCCTGTACT CCTGCAGGCC GATCCCTTCC TGGGACTGGC CATGGCCAAC1200
 TCACTTCTGA ACCCCATCAT CTACACGCTC ACCAACCGCG ACCTGCGCCA CGCGCTCCTG1260
 10 CGCCTGGTCT GCTGCGGACG CCACTCCTGC GGCAGAGACC CGAGTGGCTC CCAGCAGTCG1320
 GCGAGCGCGG CTGAGGCTTC CGGGGGCCTG CGCCGCTGCC TGCCCCCGGG CCTTGATGGG1380
 AGCTTCAGCG GCTCGGAGCG CTCATGCC CAGCGCGACG GGCTGGACAC CAGCGGCTCC1440
 ACAGGCAGCC CCGGTGCACC CACAGCCGCC CGGACTCTGG TATCAGAACC GGCTGCAGAC1500
 TGA 1503

15 (31) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 500 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 20 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

	Met	Glu	Arg	Pro	Trp	Glu	Asp	Ser	Pro	Gly	Pro	Glu	Gly	Ala	Ala	Glu
	1				5					10				15		
25	Gly	Ser	Pro	Val	Pro	Val	Ala	Ala	Gly	Ala	Arg	Ser	Gly	Ala	Ala	Ala
				20					25				30			
	Ser	Gly	Thr	Gly	Trp	Gln	Pro	Trp	Ala	Glu	Cys	Pro	Gly	Pro	Lys	Gly
					35			40				45				
30	Arg	Gly	Gln	Leu	Leu	Ala	Thr	Ala	Gly	Pro	Leu	Arg	Arg	Trp	Pro	Ala
				50				55			60					
	Pro	Ser	Pro	Ala	Ser	Ser	Pro	Ala	Pro	Gly	Ala	Ala	Ser	Ala	His	
				65			70		75			80				

- 38 -

Ser Val Gln Gly Ser Ala Thr Ala Gly Gly Ala Arg Pro Gly Arg Arg
 85 90 95

Pro Trp Gly Ala Arg Pro Met Glu Ser Gly Leu Leu Arg Pro Ala Pro
 100 105 110

5 Val Ser Glu Val Ile Val Leu His Tyr Asn Tyr Thr Gly Lys Leu Arg
 115 120 125

Gly Ala Ser Tyr Gln Pro Gly Ala Gly Leu Arg Ala Asp Ala Val Val
 130 135 140

10 Cys Leu Ala Val Cys Ala Phe Ile Val Leu Glu Asn Leu Ala Val Leu
 145 150 155 160

Leu Val Leu Gly Arg His Pro Arg Phe His Ala Pro Met Phe Leu Leu
 165 170 175

Leu Gly Ser Leu Thr Leu Ser Asp Leu Leu Ala Gly Ala Ala Tyr Ala
 180 185 190

15 Ala Asn Ile Leu Leu Ser Gly Pro Leu Thr Leu Lys Leu Ser Pro Ala
 195 200 205

Leu Trp Phe Ala Arg Glu Gly Gly Val Phe Val Ala Leu Thr Ala Ser
 210 215 220

20 Val Leu Ser Leu Leu Ala Ile Ala Leu Glu Arg Ser Leu Thr Met Ala
 225 230 235 240

Arg Arg Gly Pro Ala Pro Val Ser Ser Arg Gly Arg Thr Leu Ala Met
 245 250 255

Ala Ala Ala Ala Trp Gly Val Ser Leu Leu Leu Gly Leu Leu Pro Ala
 260 265 270

25 Leu Gly Trp Asn Cys Leu Gly Arg Leu Asp Ala Cys Ser Thr Val Leu
 275 280 285

Pro Leu Tyr Ala Lys Ala Tyr Val Leu Phe Cys Val Leu Ala Phe Val
 290 295 300

30 Gly Ile Leu Ala Ala Ile Cys Ala Leu Tyr Ala Arg Ile Tyr Cys Gln
 305 310 315 320

Val Arg Ala Asn Ala Arg Arg Leu Pro Ala Arg Pro Gly Thr Ala Gly
 325 330 335

Thr Thr Ser Thr Arg Ala Arg Arg Lys Pro Arg Ser Leu Ala Leu Leu
 340 345 350

35 Arg Thr Leu Ser Val Val Leu Leu Ala Phe Val Ala Cys Trp Gly Pro
 355 360 365

Leu Phe Leu Leu Leu Leu Asp Val Ala Cys Pro Ala Arg Thr Cys

- 39 -

	370	375	380
	Pro Val Leu Leu Gln Ala Asp Pro Phe Leu Gly Leu Ala Met Ala Asn		
	385	390	395
			400
	Ser Leu Leu Asn Pro Ile Ile Tyr Thr Leu Thr Asn Arg Asp Leu Arg		
5		405	410
			415
	His Ala Leu Leu Arg Leu Val Cys Cys Gly Arg His Ser Cys Gly Arg		
		420	425
			430
	Asp Pro Ser Gly Ser Gln Gln Ser Ala Ser Ala Ala Glu Ala Ser Gly		
		435	440
			445
10	Gly Leu Arg Arg Cys Leu Pro Pro Gly Leu Asp Gly Ser Phe Ser Gly		
		450	455
			460
	Ser Glu Arg Ser Ser Pro Gln Arg Asp Gly Leu Asp Thr Ser Gly Ser		
		465	470
			475
			480
15	Thr Gly Ser Pro Gly Ala Pro Thr Ala Ala Arg Thr Leu Val Ser Glu		
		485	490
			495
	Pro Ala Ala Asp		
		500	

(32) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 1029 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

ATGCAAGCCG TCGACAATCT CACCTCTGCG CCTGGGAACA CCAGTCTGTG CACCAGAGAC 60
 TACAAAATCA CCCAGGT CCT CTTCCCCTG CTCTACACTG TCCTGTTTT TGTTGGACTT 120
 ATCACAAATG GCCTGGCGAT GAGGATTTTC TTTCAAATCC GGAGTAAATC AAACTTATT 180
 ATTTTTCTTA AGAACACAGT CATTCTGAT CTTCTCATGA TTCTGACTTT TCCATTCAA 240
 30 ATTCTTAGTG ATGCCAAACT GGGAACAGGA CCACTGAGAA CTTTGTTGTG TCAAGTTACC 300
 TCCGTATAT TTTATTCAC AATGTATATC AGTATTCAT TCCTGGGACT GATAACTATC 360
 GATCGCTACC AGAAGACCAC CAGGCCATT AAAACATCCA ACCCCAAAAA TCTCTGGGG 420
 GCTAAGATTC TCTCTGTTGT CATCTGGCA TTCATGTTCT TACTCTCTT GCCTAACATG 480

- 40 -

ATTCTGACCA ACAGGCAGCC GAGAGACAAG AATGTGAAGA AATGCTCTT CCTTAAATCA 540
 GAGTTCGGTC TAGTCTGGCA TGAAATAGTA AATTACATCT GTCAAGTCAT TTTCTGGATT 600
 AATTCTTAA TTGTTATTGT ATGTTATACA CTCATTACAA AAGAACTGTA CCGGTACAC 660
 GTAAGAACGA GGGGTGTAAG TAAAGTCCCC AGGAAAAAGG TGAACGTCAA AGTTTCATT 720
 5 ATCATTGCTG TATTCTTAT TTGTTTGTT CCTTCCATT TTGCCGAAT TCCTTACACC 780
 CTGAGCCAAA CCCGGGATGT CTTGACTGC ACTGCTGAAA ATACTCTGTT CTATGTGAAA 840
 GAGAGCACTC TGTGGTTAAC TTCCCTAAAT GCATGCCTGG ATCCGTTCAT CTATTTTTC 900
 CTTGCAAGT CCTTCAGAAA TTCCTTGATA AGTATGCTGA AGTGCCCAA TTCTGCAACA 960
 TCTCTGTCCC AGGACAATAG GAAAAAAGAA CAGGATGGTG GTGACCCAAA TGAAGAGACT1020
 10 CCAATGTAA

1029

(33) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 342 amino acids
 (B) TYPE: amino acid
 15 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant.
 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

20 Met Gln Ala Val Asp Asn Leu Thr Ser Ala Pro Gly Asn Thr Ser Leu
 1 5 10 15
 Cys Thr Arg Asp Tyr Lys Ile Thr Gln Val Leu Phe Pro Leu Leu Tyr
 20 25 30
 Thr Val Leu Phe Phe Val Gly Leu Ile Thr Asn Gly Leu Ala Met Arg
 35 40 45
 25 Ile Phe Phe Gln Ile Arg Ser Lys Ser Asn Phe Ile Ile Phe Leu Lys
 50 55 60
 Asn Thr Val Ile Ser Asp Leu Leu Met Ile Leu Thr Phe Pro Phe Lys
 65 70 75 80
 30 Ile Leu Ser Asp Ala Lys Leu Gly Thr Gly Pro Leu Arg Thr Phe Val
 85 90 95
 Cys Gln Val Thr Ser Val Ile Phe Tyr Phe Thr Met Tyr Ile Ser Ile
 100 105 110
 Ser Phe Leu Gly Leu Ile Thr Ile Asp Arg Tyr Gln Lys Thr Thr Arg

- 41 -

	115	120	125	
	Pro Phe Lys Thr Ser Asn Pro Lys Asn Leu Leu Gly Ala Lys Ile Leu			
	130	135	140	
	Ser Val Val Ile Trp Ala Phe Met Phe Leu Leu Ser Leu Pro Asn Met			
5	145	150	155	160
	Ile Leu Thr Asn Arg Gln Pro Arg Asp Lys Asn Val Lys Lys Cys Ser			
	165	170	175	
	Phe Leu Lys Ser Glu Phe Gly Leu Val Trp His Glu Ile Val Asn Tyr			
	180	185	190	
10	Ile Cys Gln Val Ile Phe Trp Ile Asn Phe Leu Ile Val Ile Val Cys			
	195	200	205	
	Tyr Thr Leu Ile Thr Lys Glu Leu Tyr Arg Ser Tyr Val Arg Thr Arg			
	210	215	220	
15	Gly Val Gly Lys Val Pro Arg Lys Lys Val Asn Val Lys Val Phe Ile			
	225	230	235	240
	Ile Ile Ala Val Phe Phe Ile Cys Phe Val Pro Phe His Phe Ala Arg			
	245	250	255	
	Ile Pro Tyr Thr Leu Ser Gln Thr Arg Asp Val Phe Asp Cys Thr Ala			
	260	265	270	
20	Glu Asn Thr Leu Phe Tyr Val Lys Glu Ser Thr Leu Trp Leu Thr Ser			
	275	280	285	
	Leu Asn Ala Cys Leu Asp Pro Phe Ile Tyr Phe Phe Leu Cys Lys Ser			
	290	295	300	
25	Phe Arg Asn Ser Leu Ile Ser Met Leu Lys Cys Pro Asn Ser Ala Thr			
	305	310	315	320
	Ser Leu Ser Gln Asp Asn Arg Lys Lys Glu Gln Asp Gly Gly Asp Pro			
	325	330	335	
	Asn Glu Glu Thr Pro Met			
	340			

30 (34) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1077 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

- 42 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

ATGTCGGTCT GCTACCGTCC CCCAGGAAAC GAGACACTGC TGAGCTGGAA GACTTCGCGG 60
 GCCACAGGCA CAGCCTTCCT GCTGCTGGCG GCGCTGCTGG GGCTGCCTGG CAACGGCTTC 120
 GTGGTGTGGA GCTTGGCGGG CTGGCGGCCT GCACGGGGGC GACCGCTGGC GGCCACGCTT 180
 5 GTGCTGCACC TGGCGCTGGC CGACGGCGCG GTGCTGCTGC TCACGCCGCT CTTTGTGGCC 240
 TTCCTGACCC GGCAGGCCTG GCCGCTGGGC CAGGCAGGCT GCAAGGCGGT GTACTACGTG 300
 TGCGCGCTCA GCATGTACGC CAGCGTGCTG CTCACCGGCC TGCTCAGCCT GCAGCGCTGC 360
 CTCGCAGTCA CCCGCCCTT CCTGGCGCCT CGGCTGCGCA GCCCGGCCCT GGCCCGCCGC 420
 CTGCTGCTGG CGGTCTGGCT GGCCGCCCTG TTGCTCGCCG TCCCGGCCGC CGTCTACCGC 480
 10 CACCTGTGGA GGGACCGCGT ATGCCAGCTG TGCCACCCGT CGCCGGTCCA CGCCGCCGCC 540
 CACCTGAGCC TGGAGACTCT GACCGCTTTC GTGCTTCCTT TCAGGGCTGAT GCTCGGCTGC 600
 TACAGCGTGA CGCTGGCACG GCTGCGGGGC GCCCGCTGGG GCTCCGGGGCG GCACGGGGCG 660
 CGGGTGGGCC GGCTGGTGAG CGCCATCGTG CTTGCCTTCG GCTTGCTCTG GGCCCCCTAC 720
 CACGCAGTCA ACCTTCTGCA GGCGGTGCGA GCGCTGGCTC CACCGGAAGG GGCTTGGCG 780
 15 AAGCTGGCG GAGCCGGCCA GGCGGCGCGA GCGGGAACTA CGGCCTTGGC CTTCTTCAGT 840
 TCTAGCGTCA ACCCGGTGCT CTACGTCTTC ACCGCTGGAG ATCTGCTGCC CGGGCAGGT 900
 CCCCCTTCC TCACGCGGCT CTTCGAAGGC TCTGGGGAGG CCCGAGGGGG CGGCCGCTCT 960
 AGGGAAAGGGA CCATGGAGCT CCGAAACTACC CCTCAGCTGA AAGTGGTGGG GCAGGGCCGC 1020
 GGCAATGGAG ACCCGGGGGG TGGGATGGAG AAGGACGGTC CGGAATGGGA CCTTTGA 1077

20 (35) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 358 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- 25 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Met Ser Val Cys Tyr Arg Pro Pro Gly Asn Glu Thr Leu Leu Ser Trp
 1 5 10 15

30 Lys Thr Ser Arg Ala Thr Gly Thr Ala Phe Leu Leu Ala Ala Leu

- 43 -

	20	25	30	
	Leu Gly Leu Pro Gly Asn Gly Phe Val Val Trp Ser	Leu Ala Gly Trp		
	35	40	45	
5	Arg Pro Ala Arg Gly Arg Pro Leu Ala Ala Thr	Leu Val Leu His	Leu	
	50	55	60	
	Ala Leu Ala Asp Gly Ala Val Leu Leu Leu Thr	Pro Leu Phe Val Ala		
	65	70	75	80
	Phe Leu Thr Arg Gln Ala Trp Pro Leu Gly Gln Ala Gly Cys Lys Ala			
	85	90	95	
10	Val Tyr Tyr Val Cys Ala Leu Ser Met Tyr Ala Ser Val	Leu Leu Thr		
	100	105	110	
	Gly Leu Leu Ser Leu Gln Arg Cys Leu Ala Val Thr Arg Pro Phe Leu			
	115	120	125	
15	Ala Pro Arg Leu Arg Ser Pro Ala Leu Ala Arg Arg Leu Leu Leu Ala			
	130	135	140	
	Val Trp Leu Ala Ala Leu Leu Ala Val Pro Ala Ala Val Tyr Arg			
	145	150	155	160
	His Leu Trp Arg Asp Arg Val Cys Gln Leu Cys His Pro Ser Pro Val			
	165	170	175	
20	His Ala Ala Ala His Leu Ser Leu Glu Thr Leu Thr Ala Phe Val Leu			
	180	185	190	
	Pro Phe Gly Leu Met Leu Gly Cys Tyr Ser Val Thr Leu Ala Arg Leu			
	195	200	205	
25	Arg Gly Ala Arg Trp Gly Ser Gly Arg His Gly Ala Arg Val Gly Arg			
	210	215	220	
	Leu Val Ser Ala Ile Val Leu Ala Phe Gly Leu Leu Trp Ala Pro Tyr			
	225	230	235	240
	His Ala Val Asn Leu Leu Gln Ala Val Ala Ala Leu Ala Pro Pro Glu			
	245	250	255	
30	Gly Ala Leu Ala Lys Leu Gly Gly Ala Gly Gln Ala Ala Arg Ala Gly			
	260	265	270	
	Thr Thr Ala Leu Ala Phe Phe Ser Ser Ser Val Asn Pro Val Leu Tyr			
	275	280	285	
35	Val Phe Thr Ala Gly Asp Leu Leu Pro Arg Ala Gly Pro Arg Phe Leu			
	290	295	300	
	Thr Arg Leu Phe Glu Gly Ser Gly Glu Ala Arg Gly Gly Arg Ser			
	305	310	315	320

- 44 -

Arg Glu Gly Thr Met Glu Leu Arg Thr Thr Pro Gln Leu Lys Val Val
 325 330 335

Gly Gln Gly Arg Gly Asn Gly Asp Pro Gly Gly Gly Met Glu Lys Asp
 340 345 350

5 Gly Pro Glu Trp Asp Leu
 355

(36) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

10 (A) LENGTH: 1005 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

15 ATGCTGGGGA TCATGGCATG GAATGCAACT TGCAAAAACT GGCTGGCAGC AGAGGCTGCC 60
 CTGGAAAAGT ACTACCTTTC CATTTTTAT GGGATTGAGT TCGTTGTGGG AGTCCTTGGG 120
 AATACCATTG TTGTTTACGG CTACATCTTC TCTCTGAAGA ACTGGAACAG CAGTAATATT 180
 TATCTCTTTA ACCTCTCTGT CTCTGACTTA GCTTTCTGT GCACCCCTCCC CATGCTGATA 240
 AGGAGTTATG CCAATGGAAA CTGGATATAT GGAGACGTGC TCTGCATAAG CAACCGATAT 300
 20 GTGCTTCATG CCAACCTCTA TACCAGCATT CTCTTCTCA CTTTTATCAG CATAGATCGA 360
 TACTTGATAA TTAAGTATCC TTTCCGAGAA CACCTCTGC AAAAGAAAGA GTTTGCTATT 420
 TTAATCTCCT TGGCCATTG GGTTTAGTA ACCTTAGAGT TACTACCCAT ACTTCCCCTT 480
 ATAAATCCTG TTATAACTGA CAATGGCACC ACCTGTAATG ATTTGCAAG TTCTGGAGAC 540
 CCCAACTACA ACCTCATTG CAGCATGTGT CTAACACTGT TGGGGTTCT TATTCTCTT 600
 25 TTTGTGATGT GTTTCTTTA TTACAAGATT GCTCTCTTCC TAAAGCAGAG GAATAGGCAG 660
 GTTGCTACTG CTCTGCCCT TGAAAAGCCT CTCAACTTGG TCATCATGGC AGTGGTAATC 720
 TTCTCTGTGC TTTTACACC CTATCACGTC ATGCGGAATG TGAGGATCGC TTCACGCCTG 780
 GGGAGTTGGA AGCAGTATCA GTGCACTCAG GTCGTCATCA ACTCCTTTA CATTGTGACA 840
 CGGCCTTGG CCTTTCTGAA CAGTGTTCATC AACCCCTGTCT TCTATTTCT TTTGGGAGAT 900
 30 CACTTCAGGG ACATGCTGAT GAATCAAATG AGACACAACT TCAAATCCCT TACATCCTT 960
 AGCAGATGGG CTCATGAACT CCTACTTCA TTCAGAGAAA AGTGA

- 45 -

(37) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 334 amino acids
- (B) TYPE: amino acid
- 5 (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

10	Met Leu Gly Ile Met Ala Trp Asn Ala Thr Cys Lys Asn Trp Leu Ala	1	5	10	15
	Ala Glu Ala Ala Leu Glu Lys Tyr Tyr Leu Ser Ile Phe Tyr Gly Ile			25	30
	Glu Phe Val Val Gly Val Leu Gly Asn Thr Ile Val Val Tyr Gly Tyr	35	40	45	
15	Ile Phe Ser Leu Lys Asn Trp Asn Ser Ser Asn Ile Tyr Leu Phe Asn	50	55	60	
	Leu Ser Val Ser Asp Leu Ala Phe Leu Cys Thr Leu Pro Met Leu Ile	65	70	75	80
20	Arg Ser Tyr Ala Asn Gly Asn Trp Ile Tyr Gly Asp Val Leu Cys Ile	85	90	95	
	Ser Asn Arg Tyr Val Leu His Ala Asn Leu Tyr Thr Ser Ile Leu Phe	100	105	110	
	Leu Thr Phe Ile Ser Ile Asp Arg Tyr Leu Ile Ile Lys Tyr Pro Phe	115	120	125	
25	Arg Glu His Leu Leu Gln Lys Lys Glu Phe Ala Ile Leu Ile Ser Leu	130	135	140	
	Ala Ile Trp Val Leu Val Thr Leu Glu Leu Leu Pro Ile Leu Pro Leu	145	150	155	160
30	Ile Asn Pro Val Ile Thr Asp Asn Gly Thr Thr Cys Asn Asp Phe Ala	165	170	175	
	Ser Ser Gly Asp Pro Asn Tyr Asn Leu Ile Tyr Ser Met Cys Leu Thr	180	185	190	
	Leu Leu Gly Phe Leu Ile Pro Leu Phe Val Met Cys Phe Phe Tyr Tyr	195	200	205	
35	Lys Ile Ala Leu Phe Leu Lys Gln Arg Asn Arg Gln Val Ala Thr Ala	210	215	220	

- 46 -

	Leu	Pro	Leu	Glu	Lys	Pro	Leu	Asn	Leu	Val	Ile	Met	Ala	Val	Val	Ile
	225						230				235					240
	Phe	Ser	Val	Leu	Phe	Thr	Pro	Tyr	His	Val	Met	Arg	Asn	Val	Arg	Ile
							245			250					255	
5	Ala	Ser	Arg	Leu	Gly	Ser	Trp	Lys	Gln	Tyr	Gln	Cys	Thr	Gln	Val	Val
							260			265					270	
	Ile	Asn	Ser	Phe	Tyr	Ile	Val	Thr	Arg	Pro	Leu	Ala	Phe	Leu	Asn	Ser
							275			280					285	
10	Val	Ile	Asn	Pro	Val	Phe	Tyr	Phe	Leu	Leu	Gly	Asp	His	Phe	Arg	Asp
							290			295					300	
	Met	Leu	Met	Asn	Gln	Leu	Arg	His	Asn	Phe	Lys	Ser	Leu	Thr	Ser	Phe
							305			310			315		320	
	Ser	Arg	Trp	Ala	His	Glu	Leu	Leu	Leu	Ser	Phe	Arg	Glu	Lys		
							325								330	

15 (38) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1296 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- 20 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

```

ATGCAGGCAG TTAACATTAC CCCGGAGCAG TTCTCTCGGC TGCTGCAGGA CCACAAACCTG 60
ACGCAGGGAGC AGTCATCGC TCTGTACCGG CTGCGACCGC TCGTCTACAC CCCAGAGCTG 120
25 CCGGGACGCG CCAAGCTGGC CCTCGTGCTC ACCGGCGTGC TCATCTTCGC CCTGGCGCTC 180
TTTGGCAATG CTCTGGTGTG CTACGTGGTG ACCCGCAGCA AGGCCATGCG CACCGTCACC 240
AACATCTTA TCTGCTCCTT GGCCTCAGT GACCTGCTCA TCACCTTCTT CTGCATTCCC 300
GTCACCATGC TCCAGAACAT TTCCGACAAAC TGGCTGGGGG GTGCTTTCAT TTGCAAGATG 360
GTGCCATTG TCCAGTCTAC CGCTGTTGTG ACAGAAATGC TCACTATGAC CTGCATTGCT 420
30 GTGGAAAGGC ACCAGGGACT TGTGCATCCT TTTAAAATGA AGTGGCAATA CACCAACCGA 480
AGGGCTTCA CAATGCTAGG TGTGGTCTGG CTGGTGGCAG TCATCGTAGG ATCACCCATG 540
TGGCACGTGC AACAACTTGA GATCAAATAT GACTTCCTAT ATGAAAAGGA ACACATCTGC 600
TGCTTAGAAG AGTGGACCAG CCCTGTGCAC CAGAAGATCT ACACCACCTT CATCCTTGTC 660

```

- 47 -

ATCCTCTTCC TCCTGCCTCT TATGGTGATG CTTATTCTGT ACAGTAAAAT TGGTTATGAA 720
 CTTTGGATAA AGAAAAGAGT TGGGGATGGT TCAGTGCTTC GAACTATTCA TGGAAAAGAA 780
 ATGTCCAAA TAGCCAGGAA GAAGAACGA GCTGTCATTA TGATGGTGAC AGTGGTGGCT 840
 CTCTTGCTG TGTGCTGGGC ACCATTCCAT GTTGTCCATA TGATGATTGA ATACAGTAAT 900
 5 TTTGAAAAGG AATATGATGA TGTCACAATC AAGATGATT TTGCTATCGT GCAAATTATT 960
 GGATTTCCTTCA ACTCCATCTG TAATCCCATT GTCTATGCAT TTATGAATGA AAACTTCAAA1020
 AAAAATGTTT TGTCTGCAGT TTGTTATTGC ATAGTAAATA AAACCTTCTC TCCAGCACAA1080
 AGGCATGGAA ATTCAAGGAAT TACAATGATG CGGAAGAAAG CAAAGTTTC CCTCAGAGAG1140
 AATCCAGTGG AGGAAACCAA AGGAGAAGCA TTCAGTGATG GCAACATTGA AGTCAAATTG1200
 10 TGTGAACAGA CAGAGGAGAA GAAAAAGCTC AAACGACATC TTGCTCTCTT TAGGTCTGAA1260
 CTGGCTGAGA ATTCTCCTTT AGACAGTGGG CATTAA 1296

(39) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
 15 (A) LENGTH: 431 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

20 Met Gln Ala Leu Asn Ile Thr Pro Glu Gln Phe Ser Arg Leu Leu Arg
 1 5 10 15

Asp His Asn Leu Thr Arg Glu Gln Phe Ile Ala Leu Tyr Arg Leu Arg
 20 25 30

25 Pro Leu Val Tyr Thr Pro Glu Leu Pro Gly Arg Ala Lys Leu Ala Leu
 35 40 45

Val Leu Thr Gly Val Leu Ile Phe Ala Leu Ala Leu Phe Gly Asn Ala
 50 55 60

Leu Val Phe Tyr Val Val Thr Arg Ser Lys Ala Met Arg Thr Val Thr
 65 70 75 80

30 Asn Ile Phe Ile Cys Ser Leu Ala Leu Ser Asp Leu Leu Ile Thr Phe
 85 90 95

Phe Cys Ile Pro Val Thr Met Leu Gln Asn Ile Ser Asp Asn Trp Leu

- 48 -

	100	105	110	
	Gly Gly Ala Phe Ile Cys Lys Met Val Pro Phe Val Gln Ser Thr Ala			
	115	120	125	
5	Val Val Thr Glu Met Leu Thr Met Thr Cys Ile Ala Val Glu Arg His			
	130	135	140	
	Gln Gly Leu Val His Pro Phe Lys Met Lys Trp Gln Tyr Thr Asn Arg			
	145	150	155	160
	Arg Ala Phe Thr Met Leu Gly Val Val Trp Leu Val Ala Val Ile Val			
	165	170	175	
10	Gly Ser Pro Met Trp His Val Gln Gln Leu Glu Ile Lys Tyr Asp Phe			
	180	185	190	
	Leu Tyr Glu Lys Glu His Ile Cys Cys Leu Glu Glu Trp Thr Ser Pro			
	195	200	205	
15	Val His Gln Lys Ile Tyr Thr Phe Ile Leu Val Ile Leu Phe Leu			
	210	215	220	
	Leu Pro Leu Met Val Met Leu Ile Leu Tyr Ser Lys Ile Gly Tyr Glu			
	225	230	235	240
	Leu Trp Ile Lys Lys Arg Val Gly Asp Gly Ser Val Leu Arg Thr Ile			
	245	250	255	
20	His Gly Lys Glu Met Ser Lys Ile Ala Arg Lys Lys Lys Arg Ala Val			
	260	265	270	
	Ile Met Met Val Thr Val Val Ala Leu Phe Ala Val Cys Trp Ala Pro			
	275	280	285	
25	Phe His Val Val His Met Met Ile Glu Tyr Ser Asn Phe Glu Lys Glu			
	290	295	300	
	Tyr Asp Asp Val Thr Ile Lys Met Ile Phe Ala Ile Val Gln Ile Ile			
	305	310	315	320
	Gly Phe Ser Asn Ser Ile Cys Asn Pro Ile Val Tyr Ala Phe Met Asn			
	325	330	335	
30	Glu Asn Phe Lys Lys Asn Val Leu Ser Ala Val Cys Tyr Cys Ile Val			
	340	345	350	
	Asn Lys Thr Phe Ser Pro Ala Gln Arg His Gly Asn Ser Gly Ile Thr			
	355	360	365	
35	Met Met Arg Lys Lys Ala Lys Phe Ser Leu Arg Glu Asn Pro Val Glu			
	370	375	380	
	Glu Thr Lys Gly Glu Ala Phe Ser Asp Gly Asn Ile Glu Val Lys Leu			
	385	390	395	400

- 49 -

Cys Glu Gln Thr Glu Glu Lys Lys Leu Lys Arg His Leu Ala Leu
405 410 415

Phe Arg Ser Glu Leu Ala Glu Asn Ser Pro Leu Asp Ser Gly His
420 425 430

5 (40) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CTGTGTACAG CAGTCGCAG AGTG

24

(41) INFORMATION FOR SEQ ID NO:40:

15 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GAGTGCCAGG CAGAGCAGGT AGAC

24

(42) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

25

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

30

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

CCCGAATTCC TGCTTGCTCC CAGCTTGGCC C

31

(43) INFORMATION FOR SEQ ID NO:42:

- 50 -

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- 5 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

TGTGGATCCT GCTGTCAAAG GTCCCATTCC GG 32

10 (44) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- 15 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

TCACAAATGCT AGGTGTGGTC 20

20 (45) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- 25 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

TGCATAGACA ATGGGATTAC AG 22

30 (46) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 511 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

- 51 -

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

TCACAATGCT AGGTGTGGTC TGGCTGGTGG CAGTCATCGT AGGATCACCC ATGTGGCACG 60
5 TGCAACAACT TGAGATCAAA TATGACTTCC TATATGAAAA GGAACACATC TGCTGCTTAG 120
AAGAGTGGAC CAGCCCTGTG CACCAGAAGA TCTACACCAC CTTCATCCTT GTCATCCTCT 180
TCCTCCTGCC TCTTATGGTG ATGCTTATTG TGTACGTAAA ATTGGTTATG AACTTTGGAT 240
AAAGAAAAGA GTTGGGGATG GTTCAGTGCT TCGAACTATT CATGGAAAAG AAATGTCCAA 300
AATAGCCAGG AAGAAGAAAC GAGCTGTCAT TATGATGGTG ACAGTGGTGG CTCTCTTGC 360
10 TGTGTGCTGG GCACCATTCC ATGTTGTCCA TATGATGATT GAATACAGTA ATTTGAAAAA 420
GGAATATGAT GATGTCACAA TCAAGATGAT TTTTGCTATC GTGCAAATTA TTGGATTTTC 480
CAACTCCATC TGTAATCCCA TTGTCTATGC A 511

(47) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

CTGCTTAGAA GAGTGGACCA G

21

(48) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

30 (iv) ANTI-SENSE: NO

- 52 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

CTGTGCACCA GAAGATCTAC AC

22

(49) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

10 (iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CAAGGATGAA GGTGGTAG A

21

(50) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

GTGTAGATCT TCTGGTGCAC AGG

23

(51) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GCAATGCAGG TCATAGTGAG C

21

(52) INFORMATION FOR SEQ ID NO:51:

- 53 -

5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

10 TGGAGCATGG TGACGGGAAT GCAGAAG

27

(53) INFORMATION FOR SEQ ID NO:52:

15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

20 GTGATGAGCA GGTCACTGAG CGCCAAG

27

(54) INFORMATION FOR SEQ ID NO:53:

25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

30 GCAATGCAGG CGCTTAACAT TAC

23

(55) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs

- 54 -

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

5 (iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

TTGGGTTACA ATCTGAAGGG CA

22

(56) INFORMATION FOR SEQ ID NO:55:

10 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

15 (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

ACTCCGTGTC CAGCAGGACT CTG

23

(57) INFORMATION FOR SEQ ID NO:56:

20 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

25 (iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

TGCGTGTTC TGGACCCCTCA CGTG

24

(58) INFORMATION FOR SEQ ID NO:57:

30 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- 55 -

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

CAGGCCTTGG ATTTAATGT CAGGGATGG

29

5 (59) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

GGAGAGTCAG CTCTGAAAGA ATTCAAGG

27

15 (60) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

TGATGTGATG CCAGATACTA ATAGCAC

27

25 (61) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

30

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

- 56 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

CCTGATTCA TTAGGTGAGA TTGAGAC

27

(62) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

10 (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GACAGGTACC TTGCCATCAA G

21

(63) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20 (iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

CTGCACAATG CCAGTGATAA GG

22

(64) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

30 (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

CTGACTTCTT GTTCCTGGCA GCAGCGG

27

- 57 -

(65) INFORMATION FOR SEQ ID NO:64:

5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

10 AGACCAGCCA GGGCACGCTG AAGAGTG

27

(66) INFORMATION FOR SEQ ID NO:65:

15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

20 GATCAAGCTT CCATCCTACT GAAACCATGG TC

32

(67) INFORMATION FOR SEQ ID NO:66:

25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

30 GATCAGATCT CAGTCCAAT ATTACACACCA CCGTC

35

(68) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:

- 58 -

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

CTGGTGTGCT CCATGGCATC CC

22

(69) INFORMATION FOR SEQ ID NO:68:

10 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

GTAAGCCTCC CAGAACGAGA GG

22

(70) INFORMATION FOR SEQ ID NO:69:

20 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

CAGCGCAGGG TGAAGCCTGA GAGC

24

(71) INFORMATION FOR SEQ ID NO:70:

30 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- 59 -

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

GGCACCTGCT GTGACCTGTG CAGG

24

5 (72) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

GTCCTGCCAC TTCTGAGACAT GG

22

15 (73) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

GAAACTTCTC TGCCCTTACC GTC

23

25 (74) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

30

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

- 60 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

CCAACACCAG CATCCATGGC ATCAAG

26

(75) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

10 (iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

GGAGAGTCAG CTCTGAAAGA ATTCAAGG

27